PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 16 September 15, 1930 Number 9

THE RÔLE OF CHEMICAL STIMULATION IN THE SPAWNING REACTIONS OF OSTREA VIRGINICA AND OSTREA GIGAS

By Paul S. Galtsoff

U. S. Bureau of Fisheries*

Read before the Academy April 28, 1930

It is a well-known fact that chemical stimulation plays an important rôle in the sexual phenomena of many invertebrates. In the majority of cases the complexity of the reaction, consisting of a series of sexual reflexes, precludes the possibility of an accurate physiological analysis. In several water animals, especially in those that possess no organs of copulation, the reaction occurs in a simpler form and, therefore, presents a better opportunity for an experimental study. In 1913 Lillie and Just demonstrated that the males of *Nereis limbata* placed in the water in which the female worm of the same species had been kept for some time immediately began to shed sperm. The presence of sperm in the water in turn incited the shedding of eggs by the female. The reaction, according to these authors, is almost mechanical in its regularity. Fox (1924) working on sea urchins found that extrusion of genital product by one individual causes spawning of all other ripe individuals in the neighborhood.

The present investigation was carried out with two species of oyster, *Ostrea virginica*, the common American species, and *Ostrea gigas*, the Japanese oyster grown in Puget Sound. Because of the character of the spawning reaction of the oyster, it was possible to carry out the work under controlled laboratory conditions and to measure both the strength of the stimulus and the effect produced by it. Oysters used in the experiments were kept in 30-liter glass tanks; the temperature of the water, its oxygen tension, pH, and salinity were kept constant. The movement of shell was recorded by means of a slow motion kymograph. Experiments with *O. virginica* were carried out during the summers of 1927–1929 at Woods Hole. All the experiments with *O. gigas* were made in October, 1929, at Hopkins Marine Station, Pacific Grove, Calif.

Spawning of the female oyster consists of a series of the following reactions: The eggs are discharged from the ovary and fill up the mantle

cavity; the edges of the mantle on the two opposite sides come together and close the cavity, leaving but a small opening; the adductor muscle undergoes a series of rhythmical contractions, causing the discharge of the eggs, which are expelled into the water through the opening between the edges of the mantle.

Churchill (1920), Nelson (1928), and Prytherch (1929) have shown that, in nature, oysters spawn when the temperature of the water has remained for some time above 20°C. Under laboratory conditions, only a few females could be induced to spawn by a rise of temperature. Out of over two hundred experiments with *O. virginica*, positive results were obtained only with two specimens that spawned after they were taken from

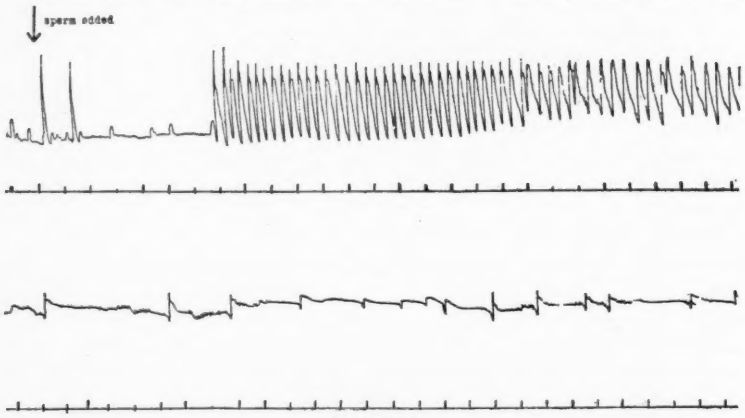


FIGURE 1

Kymograph record of the spawning reaction of *Ostrea virginica*. Lower line represents normal shell movement of the same specimen. Time intervals—one minute.

water of 19° and kept for several hours at a temperature of 27.5°. Spawning of *O. gigas* could be very easily induced by raising the temperature to 30°. In that case the response was immediate.

In both species spawning can be provoked by the addition of a small amount of sperm to the water; a typical case of the reaction is shown in figure 1. In several instances when the oyster was not ripe or after it had already spawned several times, the contractions of the adductor muscle were not accompanied by the discharge of eggs. This indicates that the adductor muscle is not stimulated by the presence of eggs in the mantle cavity or by their passage from the ovary to the cavity. No closing up of the mantle edges was observed when eggs were not discharged.

The latent period of the reaction in *Ostrea virginica* varied from 6 to 38 minutes and, as can be seen by examining table 1, was independent

of the concentration of sperm added. The coefficient of correlation computed from the data presented in table 1 is $r = \pm 0.093 \pm 0.106$, indicating that no significant correlation existed between the two variables. In *O. gigas* the latent period varied from $9^1/_2$ to 17 minutes.

TABLE 1

SHOWING NO SIGNIFICANT CORRELATION BETWEEN THE CC. OF SPERM ADDED AND THE DURATION OF THE LATENT PERIOD

LATENT PERIOD, IN MINUTES	CC. OF SPERM ADDED 0–1.9	2.0–3.9	4.0–5.9	8.0–10.0	TOTAL
35.0–40.0	1	1	1	1	4
30.0–34.9	1	1	1		3
25.0–29.9		1	1		2
20.0–24.9	1	2	2		5
15.0–19.9	2		1	1	4
10.0–14.9	7	3	2		12
5.0– 9.9	5	2	3		10
Totals	17	10	11	2	40

The duration of the reaction varied from $3^1/_2$ to 74 minutes in *Ostrea virginica* and from $13^1/_2$ to 64 in *O. gigas*. No correlation was found either between the concentration of sperm and the duration of the reaction, or between the latent period and the duration. Spawning reaction appears to follow the "All or none principle." It can be initiated by a minimum dose of sperm, which was found to be equal to 0.03 cc. of standard sperm suspension (1 gm. of gonad tissue in 100 cc. of sea water) added to 30 liters of water. This represents a concentration of approximately 150 spermatozoa in each cubic centimeter of water.

Temperature of 20° or above is prerequisite for the spawning of *Ostrea virginica;* if the water is colder than 20°, the sperm fails to provoke the reaction. For *O. gigas*, the temperature limit is 25°.

An analysis of the kymograph tracings shows that so far as the contractions of the adductor muscle are concerned, there is no difference whether the reaction has been initiated by the temperature or by the sperm. The inference can be made that the active principle of sperm suspension does not directly affect the neuro-muscular mechanism but either initiates a chain of reactions which terminate in the contractions of the adductor muscle or causes a release from the tissues of the organism of some substance which affects the neuro-muscular mechanism in a peculiar way. The action of sperm is entirely different from the action of adrenalin, which also causes rhythmical contractions of the adductor muscle. In the latter case, besides the difference in the character of the curve, the latent period lasts only a few seconds.

The spawning reaction is apparently specific, because no positive response was obtained when the sperm of *Mytilus* or *Mya* was used. Control

experiments with the same specimens of oysters gave positive reactions. Unfortunately, opportunity did not present itself to test the effect of sperm of *O. gigas* on *O. virginica* and vice versa.

The spawning reaction of the female is followed by the refractory period during which the organism is insusceptible to sperm. In *Ostrea virginica* this period lasts from 2 to 5 days.

In order to determine whether the active principle is located in the spermatozoa or is present in solution, sperm suspensions were filtered through collodion membranes. In all the experiments, the filtrate was inactive, while the residue contained the active principle. These experiments prove that the active principle is either insoluble in the sea water or is of a large molecular size and fails to pass through the membrane.

The active principle is labile; it is destroyed by heating for fifteen minutes at 55°C.

The spawning reaction of the male consists in shedding of sperm, which is discharged through the cloaca and is carried away with the outgoing stream of water. The adductor muscle is not involved in the reaction, and the shells of the spawning males remaining wide open all the time. In some of the males, shedding of sperm can be provoked by the rise of temperature to 22–23°. Nearly all of them spawn when the temperature is increased to 26–27°. The reaction can be induced also by the addition of egg suspension, the males responding almost immediately by the discharge of sperm. The latent period of the reaction varies from 5 to 12 seconds. The reaction is not followed by a refractory period and can be repeated a number of times until the male is spent.

The active principle of egg suspension is soluble in the sea water and withstands boiling for 10 minutes.

The difference in the latent periods of the spawning reactions of the female and male oysters is apparently due to the fact that the active principle of egg suspension, being soluble in the sea water, directly affects the receptors of the male; whereas the active principle of sperm suspension, because of its insolubility in water or on account of its failure to penetrate through the cell walls, must be first absorbed by the organism and affects the female through the circulatory system. Inasmuch as it has been shown by Yonge (1926–28) that absorption in the oyster takes place only through the digestive tract, one may assume that the sperm affects the female after it has reached its stomach and digestive gland. An indirect evidence corroborating this suggestion is found in the fact that particles suspended in water and caught by the gill epithelium reach the stomach in not less than 5 minutes. That the sperm may act upon the organism through the digestive tract has been demonstrated by Just (1914), who found that shedding of eggs by *Platynereis megalops* is induced by the sperm which had been introduced into its mouth by the male.

Male oysters are not only stimulated by egg suspensions but respond also to the presence of sperm in the water. In several instances ripe males attached to the kymograph were kept for two or three hours in water at 24° but failed to spawn. The discharge of sperm was, however, provoked by the addition of sperm. In that case, the latent period of the reaction, similar to the case of a spawning reaction of a female, varied from 9 to 21 minutes.

Mutual stimulation of the two sexes plays an important rôle in the propagation of oysters. The males respond more readily than the females to the increase in temperature and, under natural conditions, are apparently the first to spawn. The reaction once started by one of the males, stimulates the females and the males nearby, which in turn stimulate the others. In that way spawning spreads all over the oyster bank. The chemical stimulation is influenced in two ways by the temperature: the latter determines the lowest limit (20° for *Ostrea virginica*, 25° for *O. gigas*), below which the females fail to spawn, and supplies the initial stimulus for the spawning of the male.

* Published by permission of the United States Commissioner of Fisheries.

Churchill, E. P., 1920, App. 8, *Report U. S. Com. Fisheries for* **1919**, pp. 1–51, Doc. No. 890.

Fox, M., 1924, *Proc. Cambr. Phil. Soc., Biol. Sc.*, **1**, 1923–25, pp. 71–74.

Galtsoff, P. S., 1928, *J. Gen. Phys.*, **11**, pp. 415–431.

Just, E. E., 1914, *Biol. Bull.*, **27**, pp. 201–212.

Lillie, F. R., and Just, E. E., 1913, *Biol. Bull.*, **24**, pp. 147–159.

Nelson, Th. C., 1928, *Ecology*, **9**, p. 145–154.

Prytherch, H. F., 1929, *Bull. Bureau Fisheries*, **44**, pp. 429–503. Doc. No. 1054.

ON THE FREE ENERGY OF GLUCOSE AND OF TRIPALMITIN

BY HENRY BORSOOK AND HOWARD M. WINEGARDEN

KERCKHOFF BIOLOGICAL LABORATORIES, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated August 7, 1930

The theoretical maximum amount of work derivable from a chemical reaction is a quantity which we may designate as the reversible work, and which is equal to the decrease in the free energy of the system plus the change in the pressure–volume product. In the form of an equation this is: $W_R = -\Delta F + \Delta(PV)$.

Since the magnitude of the change in the pressure–volume product is rarely significant compared to the free energy change, for practical purposes we may consider the theoretical maximum amount of work derivable from the chemical reactions discussed here as a quantity equal to the decrease in

free energy. It is possible now-a-days to compute the free energy change for most reactions involving only inorganic reactants with a high degree of precision. On the other hand, until quite recently the standard free energy data for organic compounds were too scanty to be useful in considerations of the energy changes in physiological reactions. Instead, tacitly it has been assumed that the change in the heat content, e.g., the heat of combustion, was equal to the theoretical maximum available work. This assumption is made, for instance, in the studies of the relative efficiencies of carbohydrate and of fat as fuels for muscular work; in the principle of replaceability of "isodynamic" quantities of different foodstuffs.

Recently Parks and his collaborators[1] have determined the entropies of a number of organic compounds among which are such elementary metabolites as glucose, palmitic acid and glycerol. From the values of the standard molal entropies and the available data on the heats of combustion the standard free energies of these compounds were computed.

This work renders it possible to compute now with some precision the energy changes incurred in the oxidation of glucose and of tripalmitin. The ratios for each of these metabolites of the reversible work as defined above to the total energy change, i.e., to the heat of combustion plus the change in the pressure-volume product, can be considered as measures of their theoretical "efficiencies" as sources or generators of energy other than heat, i.e., as fuels for work, whether mechanical, chemical, or electrical.

The conventional term "efficiency" is not used with exactly the same meaning always. In the physiological literature it is employed in the sense of the ratio of work performed by the animal to the heats of combustion of the metabolites consumed in the performance of this work. In accordance with this usage we have considered the theoretical maximum "efficiency" of a physiological fuel as the ratio of the reversible work to the total energy change. When "efficiency" is defined in this way the ratio may in some cases, as Burk has shown,[2] be greater than 1. Here, obviously the balance sheet is incomplete: some reactions have been omitted, or the process is an endothermic one when it proceeds under reversible conditions, drawing heat from the surroundings. "Efficiency" might also be defined in the sense of the ratio of the work performed to the calculated maximum work obtainable from the reactions involved. For clarity the term "efficiency" will be defined here explicitly in each case. The available data are sufficient also for the calculation of the free energy change involved in the conversion of tripalmitin to glucose, or vice versa.

The usefulness of these calculations is limited before all else, of course, to those systems in which the second law of thermodynamics is obeyed. In spite of the hesitancy prevalent,[3,4] there does not appear to be any strong *a priori* reason for rejecting the conclusions of thermodynamic calculations when the available data are as adequate as they are now, in such macro-

scopic physiological chemical changes as the oxidation of glucose or of tripalmitin to carbon dioxide and water. The accuracy of the data depends now only on the accuracy of the values for the heats of combustion, and to a lesser degree on the values employed for the concentrations of the reactants in these physiological reactions.

The results of the calculations given in detail below are that the reversible work, which is practically the same as the free energy change, and which is equivalent to the theoretical maximum obtainable work in the physiological combustion of glucose, is 101 per cent of the heat of combustion; and of tripalmitin 98 per cent of the heat of combustion. This difference between glucose and a saturated fat represents, therefore, their relative theoretical "efficiencies" as fuels. In the case of the oxidation of carbohydrate the pressure-volume product change is practically zero, as the respiratory quotient for the reaction is 1. In the case of the oxidation of tripalmitin the change in the pressure-volume product is equivalent approximately to 13,000 calories. Even this quantity is negligible in comparison with the values for the heat of combustion and the free energy change which are both over 7,000,000 calories per gram mol. The computations show also that the production of fat from glucose is a process in which there is a large gain in free energy. From the assumption made here, that the second law of thermodynamics is obeyed in the body, it follows that this process is not a spontaneous one: i.e., when it does occur it is effected at the expense of the energy derived from some other reaction.

This is shown in the feeding experiments of Bleibtreu,[5] Rapport, Weiss and Csonka,[6] and Wierzuchowski and Ling.[7] Bleibtreu pointed out that the high respiratory quotients which are obtained when carbohydrate is converted to fat are most frequently accompanied, not by a decreased, but by an increased consumption of oxygen. In the experiments of Wierzuchowski and Ling in which a respiratory quotient as high as 1.55 was obtained, the oxygen consumption was nearly trebled, while the CO_2 production was quadrupled. It was calculated that, over a period of 24 hours, fat equivalent in calories to 184 per cent of the basal was manufactured from carbohydrate and stored by the animal, while the increase in metabolism over the same period was 60 per cent of the basal. Their most striking experiment was "When starch and glucose were given together on the morning of a day following a large ingestion of starch, the metabolism rose 100 per cent above the basal, and the production of fat from carbohydrate amounted to a storage of caloric energy in the form of fat equal to 260 per cent of the basal metabolism." Similarly Rapport, Weiss and Csonka observed in the hog that, when a quantity of fat whose caloric equivalent was equal to 84 per cent of the basal, was produced from carbohydrate, there was an increase in metabolism over the basal of 22 per cent. The evidence is not conclusive because it is not possible from this data to sepa-

rate the specific dynamic action due to fat production, from that due simply to the ingestion of large amounts of carbohydrate, i.e., to the effect which Lusk has designated as the specific dynamic action due to plethora. The latter effect, however, subsides within 4 or 5 hours after the ingestion of the carbohydrate and has never been found to provoke an increase in metabolism of the order of magnitude of 100 per cent of the basal. In one experiment Wierzuchowski and Ling observed, 20 hours after the ingestion of 700 gm. of corn starch, a respiratory quotient of 1.4 and an increase in the metabolism of 45 per cent over the basal. At this time the fat production from carbohydrate was 4.1 grams per hour. The power of the dog to transform carbohydrate to fat is apparently much less than that of the hog, and the increase in metabolism following the administration of large amounts of carbohydrate apparently is less, and is over sooner than in the case of the hog. Here also the evidence is not conclusive because there are no experiments on record in which as large amounts of sugar were given to dogs as were provided in the diets of the hogs in the experiments of Wierzuchowski and Ling. The experiments of Boyd, Hines and Leese, quoted by Lusk[8] suggest, however, that the specific dynamic action of carbohydrate in dogs would not under any circumstances approach the high figures observed in hogs. In these experiments glucose was administered continuously intravenously to a dog in such amounts that the blood sugar rose from 0.1 per cent to 0.7 per cent. Under these conditions "the maximal increase in metabolism ever recorded in a dog after glucose administration was obtained." During the injection the increase over the basal was 48 per cent. One hour after the cessation of injection the blood sugar had returned to a normal level, and the metabolic rate was now only 16 per cent above the basal. Three hours after the injection the increase over the basal was only 5 per cent. The mere deposition of carbohydrate in the tissues, therefore, cannot, it seems, account for the very high values for the specific dynamic action of carbohydrate observed in the hog many hours after ingestion of the carbohydrate.

The difference between these two animals appears to be that in the hog much more carbohydrate is converted to fat. The source of the energy for this conversion under the conditions of these experiments is derived largely from the oxidation of carbohydrate. The very large values for the specific dynamic action indicate that only a fraction of the energy released by this oxidation is actually used in the conversion; the remainder escapes unused as heat.

In the equation of Bleibtreu which was employed by Rapport, Weiss and Csonka, and by Wierzuchowski and Ling for the comparison of the values obtained from indirect with those from direct calorimetry, the specific dynamic effect of the process of fat-production from carbohydrate is tacitly implied. Bleibtreu considered that in the formation of 100 grams

of pork fat with an elementary composition of C, 76.54%, H, 11.94%, and O, 11.52%, 191.35 grams of glucose would be required in order to provide the carbon. The hydrogen of the glucose, not used to form fat, combines with some of the unused oxygen to form water; and the remaining oxygen is consumed in the combustion of more glucose, over and above the basal requirement. If the process of fat formation did not exert a specific dynamic action, this unused oxygen presumably would be available for purposes of the basal metabolism, so that, under these conditions, the consumption of inspired oxygen would be reduced. But, as Bleibtreu pointed out, the oxygen consumption is increased, in spite of the release of a large quantity of oxygen from the carbohydrate. The observations on plants of Terroine and Bonnet and their collaborators[20,21,22] show similarly that the formation of fat from carbohydrate involves an expenditure of energy which manifests itself in an addition to the basal metabolism. Their results are discussed in detail below.

It seems improbable that the production of fat from carbohydrate is carried on with the same expenditure of energy in all animals. The specific dynamic action of this process in any animal can be considered as the reciprocal of the ability of that animal to transform carbohydrate into fat. The final equation of Bleibtreu is

$$270.06 \text{ gm. glucose} = 100 \text{ gm. fat} + 115.45 \text{ gm. } CO_2 + 54.6 \text{ gm. } H_2O$$

In animals such as the hog or goose, that transform carbohydrate into fat very readily, the amount of glucose used and CO_2 and water formed to produce 100 grams of fat will be less than in a dog. The procedure of Lusk and his pupils, of first deducting the specific dynamic action, as indicated by the increased oxygen consumption, and then estimating the calories equivalent to the CO_2 production over and above that required for an R. Q. of 1, tends to obscure the differences between animals, and by an apparent more or less satisfactory agreement between direct and indirect calorimetry endows the empirical equation of Bleibtreu with an unwarranted generality.

The converse reaction, the hypothetical conversion of tripalmitin to glucose, is accompanied by a work content decrease which amounts to 21 per cent of the theoretical maximum amount of work available, and to 20 per cent of the total energy change in the combustion of tripalmitin. If fat is converted into carbohydrate before its energy can be used in muscular work, the energy equivalent to the decrease in work content must be wholly dissipated as heat instead of used for work. Under these conditions fat must necessarily be less efficient as a source of energy for work than carbohydrate.

There are described in the literature a number of attempts to determine experimentally the relative efficiencies of carbohydrate and of fat as fuels for

muscular work in the animal body. The careful study by Krogh and Lindhard of about ten years ago[9] led to a general acceptance of a figure for the efficiency of fat as a fuel for muscular work 11 per cent less than that of carbohydrate. Unfortunately the reliability of this estimate has been impaired by the subsequent work of Hill, Long and Lupton[10] on the oxygen debt, and by the studies of Rapport and Ralli,[11] and particularly of Best, Furusawa and Ridout[12] on the respiratory quotient of the excess metabolism during the performance of and recovery from muscular work. The serious deficiency in the data of Krogh and Lindhard is their omission of the metabolism during recovery. Their estimate of 11 per cent difference in efficiency between the two fuels is based upon observations on the "excess" metabolism only during the performance of the work.

The later observations of Rapport and Ralli on the "excess" metabolism of mild exercise by dogs are in accordance with those of Krogh and Lindhard in so far as they found that according to whether the diet is preponderantly carbohydrate or fat, respectively more carbohydrate or fat is burned. It is possible, therefore, that in the experiments of Krogh and Lindhard, the respiratory quotient during recovery was the same as that during the exercise. If this was the case, and if the recovery period is of the same length on the two diets, then the same relative differences that were observed on the carbohydrate and fat diets might have been found for the total excess metabolism.

Leathes and Raper[13] calculated that a quantity of heat equivalent to 20 per cent of the heat of combustion of fat is evolved in the conversion of fat to carbohydrate. They consider that this figure is at variance with the figure of 11 per cent lower efficiency of fat obtained by Krogh and Lindhard, if it is considered that fat is converted to carbohydrate before it is burned. This difference between 20 per cent and 11 per cent constitutes a discrepancy only if it is assumed that before fat can provide energy for any kind of work in any tissue of the body it must first be converted to carbohydrate. If the energy available for work released by the combustion of fat can be utilized by the viscera, without its previous conversion to carbohydrate, then even if such a conversion is obligatory in skeletal muscle, it would still be possible to reconcile the difference between the calculated 20 per cent and the observed 11 per cent lower efficiency of fat, by postulating a selective utilization of fat by the viscera and of carbohydrates by the muscles when muscular work is carried out on a high fat diet; or by some such mechanism as that suggested by Rapport and Ralli, that the energy of combustion of fuels other than a fraction of the free lactic acid may serve in the re-synthesis of glycogen in the recovery phase of muscular contraction.

But until the observations of Best, Furusawa and Ridout on the excess metabolism during muscular exercise have been explained, any conclusions

regarding the nature of the fuel used in muscular exercise can only be tentative so long as they are based, as they are above, upon inferences drawn from respiratory quotients. Best, Furusawa and Ridout found that "The respiratory quotient of the excess metabolism of muscular exercise is not invariably unity, but increases with the severity of the exercise from a value approximately the same as the basal quotient for very mild exercise, through unity for moderate exercise, to one considerably above unity for very severe exertion." These observations in so far as they pertain to excess metabolism during mild exercise confirm the similar earlier findings of Rapport and Ralli.

If eventually it is found that muscular work is performed on a fat diet with an efficiency considerably less (more than a few per cent less) than on a carbohydrate diet, this inefficiency cannot be ascribed to differences between the free energy and heat energy of these two metabolites; but rather to differences in their intermediary metabolism, whereby more energy escapes as heat in the case of fat than in the case of carbohydrate.

These conclusions regarding the losses of energy incurred both in the conversion of fat to carbohydrate and vice versa are supported by the observations of Terroine and Bonnet and their collaborators on germinating seeds and moulds. The problem is simplified with these forms on account of the absence of any specific dynamic action due to plethora.[20] In the mould Sterigmatocystis nigra Terroine and Bonnet[21] found that for every 100 calories of potential energy contained in the glucose used in the process of conversion to fat, 11 per cent was lost as heat when the fatty acid content of the mycelium rose from 3.1 per cent to 9.0 per cent; 13 per cent was lost when the fatty acid content was 10.5 per cent; and 17 per cent was lost when the fatty acid content was 12.3 per cent. If in the feeding experiments on the hog one ascribes the whole of the specific dynamic action of the carbohydrate to the work of converting carbohydrate to fat, the ratio of energy lost to the total employed is 18 per cent in the experiment of Rapport, Weiss and Csonka, and 23 per cent in the experiment of Wierzuchowski and Ling. The latter estimates are certainly too high, as only a fraction of the total specific dynamic action of carbohydrate in these feeding experiments could have arisen from the production of fat. There must have been another fraction due simply to the plethora effect. The "specific dynamic action" of formation of fat from carbohydrate will therefore be apparently the same in plants as in the hog.

The problem of the energy losses incurred in the utilization of fat in animals also is elucidated by the observations made on plants, where the production of carbohydrate from fat is unequivocal and easily demonstrated. In germinating grains Terroine, Trautmann and Bonnet[22] observed that when fat is converted into cellulose, 23 per cent of the energy of the fat used is lost. Similarly in the mould Sterigmatocystis nigra the

amount of energy lost in the conversion to carbohydrate varied, depending upon the fat used between 18.8 per cent and 25.8 per cent. These values are in accord with that calculated, and substantiate the criticism, based upon calculated energy changes, of the view that fat must first be converted to carbohydrate before it can be utilized in muscular work.

Calculations: The symbols employed are those of Lewis and Randall.[14] Throughout, the activity of the reactants will be considered as equal to their molar concentrations. This is permissible because even the allowances for the free energy of solution and dilution, and for temperature, are very small compared with the free energy changes when the reactions are considered as occurring with the metabolites and end products in their standard states. In the re-calculation of the standard free energies at 37°C. instead of at 25°C., the values of the heats of formation and of combustion will be taken to be the same at the two temperatures. The free energy values employed in the calculations which follow are taken from the table of revised values of Parks, Kelley and Huffman.

Glucose:

ΔF_{298} (solid) $= -219{,}000$ calories.

From the second-law free energy equation $\frac{-\Delta F_2}{T_2} - \frac{-\Delta F_1}{T_1} = \int_{T_1}^{T_2} \frac{\Delta H}{T_2} dT$; and $\Delta H_{298} = -303{,}000$ calories, $\Delta F_{310} = -215{,}600$ calories. Interpolating from the data in the International Critical Tables, the solubility of glucose at 37°C. is 146 gm. $C_6H_{12}O_6.H_2O$ in 100 gm. water. The mol fraction of the saturated solution at 37°C. is $\frac{\frac{132.7}{180}}{\frac{132.7}{180} + \frac{113.3}{18}} = .11$. Considering the concentration of glucose at the site of oxidation to be nearly the same as it is in the blood, normally approximately 0.1 gm. in 100 cc., the mol fraction is $\frac{\frac{0.1}{180}}{\frac{90}{18}} = .00011$, and the free energy of the dilution per mol is therefore $-\Delta F = RT \ln \frac{.11}{.00011} = 4250$ cal. $\therefore \Delta F$ glucose (mol fraction $= .00011$) $= -219{,}900$ calories.

The concentrations of oxygen, carbon dioxide and water in the blood and tissues are not those of their standard states, and accordingly slight corrections are introduced in the free energy computations for these deviations, in addition to the modification of the standard free energy values for the change in temperature from 25°C. to 37°C.

Oxygen: Krogh[15] gives the following values for the difference in oxygen tension between the blood in the capillaries and the muscles under the following physiological conditions: rest, 6.5 − 3.5 per cent of an atmosphere; massage, 2.5 per cent; work, 0.4 per cent; maximum circulation, 0.25 per cent. Considering, therefore, the oxygen tension in the muscles during work to be 100 mm. mercury, the free energy decrease per mol of oxygen due to "dilution" is $-\Delta F = RT\ln\frac{760}{100} = 1250$ calories.

H_2O:

$$\Delta F_{298} = -56560.$$
$$\Delta H = -68270.$$
$$\therefore \Delta F_{310} = -56100.$$

In plasma the mol fraction of water is .995.[16] The correction for dilution is therefore negligible.

CO_2:

$$\Delta F_{298}\ CO_2\ (g) = -94{,}260 \text{ calories.}$$
$$\Delta H = -94{,}300 \text{ calories.}$$
$$\therefore \Delta F_{310}\ CO_2\ (g) = -94{,}260 \text{ calories.}$$

Considering the pressure of CO_2 in the tissues to correspond to 40 mm. of mercury, the correction for the free energy decrease due to the dilution of the CO_2 is

$$-\Delta F = RT\ln\frac{760}{40} \text{ calories} = 1800 \text{ calories.}$$

$$\Delta F_{310}\ CO_2\ (g)\ (40 \text{ mm. Hg}) = -96{,}100 \text{ calories.}$$

Oxidation of Glucose: $C_6H_{12}O_6$ (mol fraction .00011) + $6O_2$ (g) (100 mm. Hg) = $6CO_2$ (g) (40 mm. Hg) + $6H_2O$ (l); $-\Delta F_{310} = (-219{,}900) + (-6 \times 1250) - (-6 \times 96{,}100) - (-6 \times 56{,}100) = 685{,}800$ calories. The heat of combustion of glucose in solution is 676,080 calories.[17]

$$\therefore \frac{\text{Free Energy of Combustion}}{\text{Heat of Combustion}} = \frac{685{,}800}{676{,}080} = 1.01.$$

The change in the pressure–volume product is negligible here.

The fact that the free energy of combustion is greater than the heat of combustion means only that if it were possible to convert into work all the free energy released by the oxidation of glucose, approximately 9000 calories would be absorbed per gram mol of glucose oxidized. Actually in the body, of course, only a small fraction, between 15 and 30 per cent of the free energy, is utilized, the remainder being dissipated as heat.

The experimentally determined values for the standard molal free energies of palmitic acid and of glycerol permit a fairly good approximate computation of the free energy of tripalmitin. In the hydrolysis and formation of a great variety of esters Berthelot and Saint-Gilles[18] found that the equilibrium constant for the reaction, acid + alcohol = ester + water, was always in the neighborhood of 4; and that the value of the equilibrium constant seemed to be independent of temperature in the range from 100°C. to 220°C. One may assume from these observations that for this type of reaction ΔF and ΔH are equal, or nearly so. The value of ΔF at 25°C. then becomes about −800 calories per linkage mol of ester. The value for the heats of combustion in the International Critical Tables, however, yield values for the heats of formation of esters with the higher acids as high as 15,000 calories. No heat data are available in regard to glycerol esters. Since it is insignificant in the final result here whichever value is chosen for ΔH we have preferred the alternative of assuming that the heat of ester formation in the case of tripalmitin is approximately 1000 calories per linkage. The free energy of formation of tripalmitin from glycerol and palmitic acid is accordingly taken as +2400 calories.

The reaction may be represented as follows: 1 glycerol + 3 palmitic acid ⟶ tripalmitin + 3 water; $-\Delta F = +2400$ calories.

Parks, Kelley and Huffman give −115,700, and −94,000 as the free energy values at 25°C. of glycerol and palmitic acid respectively. The free energy of water at 25°C. is −56,560. The value for the free energy of palmitic acid is based upon a value for the heat of combustion of 2,379,000 calories. The value given in the International Critical Tables is slightly higher, 2,380,000 calories. With this value, the free energy of palmitic acid is −93,000 calories. From these figures an approximate value for the free energy of tripalmitin is −227,000 calories.

Assuming that the heat of ester formation of tripalmitin is +3000 calories, ΔH, the heat of formation of tripalmitin, is simply the sum of the heats of formation of 1 gram mol of glycerol and three of palmitic acid plus 3000 calories, minus the heat of formation of three gram mols of water, and is equal to −610,500 calories. The free energy, ΔF, at 37°C. then is −211,000 calories.

Because fat in the tissues probably exists as a saturated solution in equilibrium with the solid phase there is no free energy of solution to be taken into account for tripalmitin.

Oxidation of Tripalmitin: $2C_{51}H_{98}O_6$ (solid) + $145O_2$(100 mm. Hg) ⟶ $102CO_2$ (40 mm. Hg) + $98H_2O$; $-\Delta F = (2 \times -211{,}000) + (145 \times -1250) - (102 \times -96{,}100) - (98 \times -56{,}100) = 14{,}696{,}750$ calories; = 7,348,000 calories per gram mol of tripalmitin. The diminution in the pressure–volume product is equivalent to 13,000 cal. per gram mol of tripalmitin oxidized.

There are no data available on the direct measurement of the heat of combustion of tripalmitin; but little error will be involved in assuming it to be equal to the sum of the heats of combustion of 1 mol of glycerol and 3 mols of palmitic acid. The value then is 7,537,000 calories, minus 3000 calories, the heat of esterification.

$$\therefore \frac{\text{Free energy of combustion}}{\text{Heat of combustion}} = \frac{7{,}348{,}000}{7{,}534{,}000} = .98; \quad \frac{\text{Reversible work}}{\text{Total energy change}} = \frac{7{,}335{,}000}{7{,}521{,}000} = .98.$$

The theoretical maximum work obtainable from the oxidation of tripalmitin is therefore relative to their respective heats of combustion, only 3 per cent less than that of glucose.

The free energy changes in the conversion of tripalmitin to glucose under physiological conditions may be considered in three stages: (1) the hydrolysis of tripalmitin, (2) the conversion of glycerol to glucose, (3) the conversion of tripalmitin to glucose.

The hydrolysis may be taken to incur a free energy change of -2400 calories.

Conversion of Glycerol to Glucose under Physiological Conditions:

ΔF_{298} glycerol $= -115{,}700$ calories.
ΔH_{298} glycerol $= -159{,}300$ calories.
$\therefore \Delta F_{310}$ glycerol $= -113{,}700$ calories.

Glycerol exists in the blood in a minute concentration. In order to assign some figure for the free energy of dilution we shall assume that in the liver, where the conversion may be said to occur, its concentration is of the order of magnitude of a mol fraction of 10^{-4}, so that the free energy of dilution is approximately 5000 calories.

$\therefore -\Delta F_{310}$ (mol fraction 10^{-4}) $= -118{,}700$ calories.

$\therefore 2C_3H_8O_3$ (mol fraction 10^{-4}) $+ O_2$ (100 mm. Hg) $\longrightarrow C_6H_{12}O_6$ (mol fraction .00011) $+ 2H_2O$; $-\Delta F = 103{,}400$ calories $= 51{,}700$ calories per mol of glycerol.

The pressure–volume product change is approximately 600 calories, and is negligible.

Conversion of Palmitic Acid to Glucose under Physiological Conditions:

ΔF_{298} palmitic acid $= -93{,}000$ calories.
ΔH_{298} palmitic acid $= -221{,}000$ calories.
$\therefore \Delta F_{310}$ palmitic acid $= -88{,}000$ calories.

Therefore, under physiological conditions, for the reaction:

$$3C_{16}H_{32}O_2 + 21O_2 \text{ (100 mm. Hg)} = 8C_6H_{12}O_6 \text{ (mol fraction .00011)};$$

$-\Delta F$ = 1,469,000 calories; pressure-volume product change, $\Delta(PV)$ = $-13{,}000$ calories. The total change in work content in the conversion under physiological conditions of tripalmitin to the maximum possible amount of glucose is 1,505,000 calories. Therefore for tripalmitin,

$$\frac{\Delta W_R \text{ conversion}}{W_R \text{ oxidation}} = \frac{1{,}505{,}000}{7{,}335{,}000} = .21;$$

and

$$\frac{\Delta W_R \text{ conversion}}{\text{Total energy change in oxidation}} = \frac{1{,}505{,}000}{7{,}521{,}000} = .20$$

The conversion of tripalmitin to glucose involves the liberation of a large fraction of the energy of the fat. The reverse reaction, the formation of one gram mol of tripalmitin from $8^1/_2$ gram mols of glucose, would incur a gain in the quantity of that energy which we have designated as reversible work, equivalent to the total energy of oxidation of approximately 2 gram mols of glucose. The reaction almost certainly does not take place in a reversible manner, and a better guess, it seems, would be that at least 4 gram mols of glucose would be required. The unused energy would escape as heat and in direct calorimetry would represent the specific dynamic action of this process.

A concise statement of the free energy relationship between carbohydrate and fat is obtained from the comparison of the free energy content per gram mol of carbon in glucose and in tripalmitin. Per gram mol of carbon the free energy content of glucose is $-36{,}500$ calories; and of tripalmitin -2060 calories. It is clear that any process by which carbohydrate is converted into fat in such a manner that all of the carbon of the one compound is re-constituted into the other involves a large gain in free energy which must be supplied by other processes. If the mechanism by which this energy is supplied is not a perfectly reversible one, a larger amount of energy will be produced than is used in the conversion of carbohydrate into fat; and the excess, lost as heat, will appear as specific dynamic action. Similarly the converse production of carbohydrate from fat is accompanied by a large decrease, i.e., liberation of free energy, which will appear, wholly or in part, as heat or work.

In plants, as the observations discussed above show, this released energy is not used for any other purpose, and escapes, presumably, as heat.

It follows from the small difference in the ratios of the reversible work obtainable to the total released in the combustion of glucose and of tripalmitin, that the explanation for any large differences observed experimentally in their respective efficiencies as fuels must be sought in their intermediary metabolism. If fat must be converted to glucose before it can serve as a fuel for muscular work, the large free energy change involved

in the conversion of fat into glucose renders it probable on theoretical grounds (i.e., the process probably is not perfectly reversible) that fat will be less efficient as a fuel. On the other hand, if the oxidation of fat *per se* can serve, as suggested by Rapport and Ralli, in the re-synthesis of lactic acid to glycogen, then there is no *a priori* reason for considering that fat will be less efficient than lactic acid. The plausibility of the suggestion of Rapport and Ralli is enhanced by the observations of Himwich, Koskoff and Nahum[14] that the liver possesses a great avidity for the lactic acid liberated in the blood-stream during exercise; and by the observation of Rapport and Ralli on dogs, confirmed by Best, Furusawa and Ridout on men, that the excess metabolism of mild exercise is that of the basal metabolism at that time.

If fat is first converted to glucose, as Best, Furusawa and Ridout maintain, then the free energy of conversion must be considered as lost for purposes of muscular work. It is interesting to compute from the calculations carried out above the minimum theoretical difference in efficiency of fat and carbohydrate under these circumstances. If the efficiency of the process of muscular contraction is taken first at 30 per cent, then 30 per cent of the energy of glucose derived from tripalmitin would be used, which corresponds to 23.1 per cent of the heat of combustion of the tripalmitin. If the muscular work is carried out at a 10 per cent efficiency then 7.7 per cent of the heat of combustion of tripalmitin is used. In each case fat is 77 per cent as efficient a fuel as carbohydrate. The experimental investigations into the difference in efficiency of fat and carbohydrate have in no case yielded figures as high as this, showing, if subsequent investigations do not yield a very different result, that fat is burned during muscular work as such; and that the efficiency of the processes by which this energy is used is not much less than is the case with glucose.

Krogh and Lindhard criticize the estimations of the energy changes in the conversion of fat to carbohydrate as follows: "We are not convinced of the validity of any of these summary methods of calculating the waste of energy incidental to the conversion of fat into sugar or any other substance. As a reason for suspicion against summary methods of calculating the loss of energy in question we would suggest the making up of 1 molecule of fat (tripalmitin) into 12 molecules of sugar by the addition of 21 molecules of CO_2 and 23 of H_2O, which would result in a *gain* of energy amounting to about 18 per cent." It is interesting to compute the free energy change for this proposed hypothetical reaction.

$$C_{51}H_{98}O_6 + 21CO_2 + 23H_2O \longrightarrow 12C_6H_{12}O_6$$

From the data given above the free energy change, $-\Delta F$ is $-881{,}000$ calories. In other words, this reaction could not occur except at the cost of other reactions. As it is improbable that a reaction such as this would

proceed under ideal, reversible conditions, we may expect instead of a gain in energy of 18 per cent, a wastage as heat of approximately 1,000,000 calories for each molecule of tripalmitin converted to glucose.

The computations carried out above show that the cycle of storage of carbohydrate by conversion to fat and its later reconversion to carbohydrate preliminary to its ultimate oxidation, is one which is carried out at the cost of a considerable amount of energy to the organism. From studies of the degree of reversibility of reactions *in vivo* it may be possible to estimate how much of this expended energy, i.e., not contained in the newly formed molecules as chemical potential energy, is, from the nature of the reactions and substances involved, inevitably wasted as heat, and how much under suitable conditions can be converted into work.

Summary—1. The theoretical maximum amount of work obtainable from the oxidation under physiological conditions of glucose, and of tripalmitin has been computed from the standard free energies and heats of combustion of these compounds.

2. It has been shown that the theoretical difference in efficiency, i.e., the ratio of the theoretical maximum work obtainable to the total energy change, is little different for glucose and for tripalmitin.

3. The process of conversion of glucose into tripalmitin probably exerts a considerable "specific dynamic action" on metabolism.

4. From the fact that the conversion of tripalmitin to glucose involves a release of energy approximately equivalent to 20 per cent of the theoretical maximum work available from the oxidation of tripalmitin, and the observation that the difference in efficiency of fat as a fuel for muscular work, as compared with carbohydrate, is much less than 20 per cent, it follows either that fat is burned as such in the provision of energy for muscular work, or that the energy released in the hypothetical conversion of fat to cabohydrate is not dissipated as heat, but is used for work; and the efficiency of utilization of this energy is little different from that in the consumption of carbohydrate.

The authors wish to thank Professor S. J. Bates for his careful reading of this paper, and for his valuable suggestions.

[1] Parks, G.S., Kelley, K. K., and Huffman, H. N., *J. Am. Chem. Soc.*, 1929, **51,** 1969.
[2] Burk, D., *Proc. Roy. Soc.*, 1929, **B104,** 153.
[3] Guye, C. E., *Physico-Chemical Evolution*, London, **1925.**
[4] Donnan, F. G., *J. Gen. Physiol.*, 1925–7, **8,** 685.
[5] Bleibtreu, M., *Archiv. ges. Physiol.*, 1901, **85,** 345.
[6] Rapport, D., Weiss, R., and Csonka, F. A., *J. Biol. Chem.*, 1924, **60,** 583.
[7] Wierzuchowski, M., and Ling, S. M., *J. Biol. Chem.*, 1925, **64,** 697.
[8] Lusk, G., *The Elements of the Science of Nutrition*, 4th ed., Philadelphia and London, **1928.**
[9] Krogh, A., and Lindhard, J., *Biochem. J.*, 1920, **14,** 290.
[10] Hill, A. V., Long, C. N. H. and Lupton, H. *Proc. Roy. Soc.*, 1924, **B96,** 455.

[11] Rapport, D., and Ralli, E. P., *Am. J. Physiol.*, 1927–8, **83**, 450.
[12] Best, C. H., Furusawa, K. and Ridout, J. H., *Proc. Roy. Soc.*, 1929, **B104**, 119.
[13] Leathes, J. B., and Raper, H. S., *The Fats*, London, **1925.**
[14] Lewis, G. N., and Randall, M., *Thermodynamics*, New York, **1923.**
[15] Krogh, A., *The Anatomy and Physiology of Capillaries*, New Haven, **1929.**
[16] Quoted from Mathews, A. P., *Physiological Chemistry*, New York, **1927.**
[17] Davis, J. G., Slater, W. K., and Smith, V., *Biochem. J.*, 1926–7, **20**, 1155.
[18] Berthelot, M., and Saint-Gilles, L. P., *Ann. Chim. Phys.*, 1863, Series 3, **68**, 225.
[19] Himwich, H. E., Koskoff, Y. D. and Nahum, L. H., *J. Biol. Chem.*, 1930, **85**, 571.
[20] Terroine, E. F., Trautmann, S., Bonnet, R. and Hée, H., *Bull. Soc. Chim. Biol.*, 1926, **8**, 584.
[21] Terroine, E. F., and Bonnet, R., *Bull. Soc. Chim. Biol.*, 1927, **9**, 588.
[22] Terroine, E. F., Trautmann, S., and Bonnet, E., *Ann. Physiol. Physicochim. Biol.*, 1926, **2**, 172.

---

*CLAUDE BERNARD'S THEORY OF NARCOSIS**

BY WILDER D. BANCROFT AND GEORGE H. RICHTER**

BAKER LABORATORY OF CHEMISTRY, CORNELL UNIVERSITY

Communicated August 11, 1930

The predominant theory of narcosis at the present time is that of Overton and Meyer, which postulates that lipoid solubility is the important factor. It is quite evident that a substance which is soluble both in lipoids and in water will pass into the cell readily; but that does not give us a theory of narcosis, unless we postulate that all substances which pass into the cells produce anesthesia—which is obviously absurd.

It is true that Meyer did make this assumption explicitly. "All chemically indifferent substances, which are soluble in fats and similar substances, must exert a narcotic action upon living protoplasm, in so far as they diffuse therein." It is not to be believed that anybody would now admit the truth of this assumption, an assumption which ignores the known differences between sensory nerves, sympathetic nerves, and motor nerves. Consequently the theory of Overton and Meyer is at best a theory of permeability and not a theory of narcosis at all.

Over sixty years ago, Claude Bernard[1] put forward the theory that anesthesia always occurred when we have reversible coagulation of the colloids of the sensory nerves. This was rejected for two reasons: that the concentrations necessary to coagulate nerve colloids (meaning proteins) were much higher than those occurring in anesthesia; and that coagulation by anesthetics was irreversible. Claude Bernard did not know enough colloid chemistry to overthrow these two criticisms and consequently his theory went into the discard temporarily.

It is true that relatively high concentrations of alcohol, for instance, are necessary to coagulate a moderately pure, negatively charged sol—forty per cent by volume in the case of copper ferrocyanide.[2] If we add nearly enough electrolyte to precipitate the copper ferrocyanide sol, we can make the amount of alcohol necessary to coagulate it as small as we please. Under these conditions, the coagulation is reversible. The two apparently serious objections to Claude Bernard's theory have therefore been eliminated.

When an acidified albumin sol is treated with sodium sulphate until it is on the verge of precipitating, one drop of the common alcohols or a small crystal of chloral hydrate is sufficient to coagulate the albumin. With a negatively charged albumin, calcium chloride can be used. Edestin acts similarly but is not sensitized so much as albumin, which we can perhaps take as typical of the colloids of the sensory nerves.

Claude Bernard merely said reversible coagulation of the nerve colloids, without committing himself whether these are proteins or lipoids. Since we know relatively little about reversible coagulation of lipoids and since some of the proteins seem to behave like the nerve colloids, it will be wise to discuss proteins only, until such time as it becomes necessary to consider the behavior of the lipoids in addition. Proteins are admittedly the important factor in the case of anaphylactic shock. For the time being, therefore, Claude Bernard's theory is that reversible coagulation of the proteins of the sensory nerves is always accompanied by anesthesia. We are making the additional explicit assumption that there is increased irritability as the nerve colloids approach the point of instability and the beginning of reversible coagulation. One would normally expect with increasing amounts of anesthetic to observe increasing irritability, anesthesia, and death in case the coagulation becomes irreversible. It is well known that there is a period of irritability before anesthesia begins and after anesthesia passes off. Since strychnine has a stimulating effect in small doses and causes death when administered in larger doses, it seemed probable that there might be a range of concentrations, perhaps very narrow, within which strychnine would act as an anesthetic. The matter is complicated by the facts that strychnine apparently affects the sympathetic nerves as well as the sensory nerves, and that it also interferes with the oxygen metabolism. In spite of this, the prediction in regard to strychnine was verified. It is a simple matter to anesthetize a frog with strychnine so that he will stay in a deep narcosis for several days. It is also possible to anesthetize a dog with strychnine provided he is given oxygen to minimize the unwanted effects of the drug.

Professor Otto Rahn of the College of Agriculture reported at a meeting of the Research Club on the action of mercuric chloride on bacteria. If the corrosive sublimate was administered in small doses or for short times,

FIGURE 1

FIGURE 2

there was no apparent deleterious action. If larger doses were given or if a small dose were allowed to act for a longer time, a range of dormancy was observed. The bacteria could be aroused from dormancy by elimination of the mercuric chloride. If the dose were still larger, a range of death was observed. This meant that no treatment would cause the bacteria to function again. Following out the analogy with ordinary anesthetics, it was predicted that the very small doses should have produced irritability and a consequent increase in the rate of growth of the colonies. This was actually observed.

Dr. M. J. Brown of the Roessler and Hasslacher Chemical Company has called our attention[3] to the effect of hydrocyanic acid on the eggs of some scale insect found on citrus trees. A relatively high concentration of hydrocyanic acid gas killed the eggs of the scale; but a low concentration stimulated the eggs so much that a larger percentage hatched than in the control groups. This is analogous to the action of corrosive sublimate on bacteria and doubtless a range of dormancy could have been found if anybody had been interested in looking for it. These are merely special illustrations of the general phenomenon that all depressor substances exert a stimulating affect when used at appropriate dilution.

Marinesco[4] has shown that dilute alcohol coagulates the colloids in living cells; but he did not show that the colloids went back when the narcotic was removed. In order to establish this for a simple case, we prepared young, vigorous cultures of ordinary baker's yeast by inoculating yeast into Laurent's medium made up with 1.5% dextrose. Subcultures were made every twenty-four hours, the third subculture being used in the experiments. The yeast cells were examined under the ultramicroscope and photographed. The culture was then treated with the narcotizing concentration of amyl alcohol, about two per cent, and again placed under the ultramicroscope. After about ten minutes the Brownian movement is perceptibly slower. A few minutes later coagulation becomes visible and then a light but very pronounced flocculation occurs. The complete change requires twenty-five minutes, although the time depends upon the concentration of the narcotic, the temperature and the individual culture. A similar coagulation was obtained by treating the yeast cells with such other narcotics as chloroform, paraldehyde, ether and chloral hydrate. Figures 1 and 2 show the yeast cell before and after flocculation.

The narcotized culture was centrifuged, the supernatant liquid poured off from the cells, and fresh, sterile medium added. The cells were washed twice in this way and examined again under the ultramicroscope. The material in the cells soon develops a slight Brownian movement, then the aggregates break up into smaller particles and the motion becomes more pronounced. At the end of twenty-five or thirty minutes, the material is peptized completely and the cell is normal in every respect. The yeast

will ferment the medium and will reproduce just like normal cells. If too much narcotic is added or if it is allowed to act for too long a time, the coagulation becomes more marked and cannot be reversed. The cell is dead. This set of experiments is a great triumph for the theory of Claude Bernard.

In terms of the theory, a local anesthetic must be a substance which is not transported rapidly by the blood and which consequently acts where it is injected. Administering adrenalin along with novocaine closes the capillaries to some extent and thereby helps localize the action of the latter.

Claude Bernard points out that both morphine and ether act first on the brain; but that morphine acts less rapidly on the sensory nerves than do the true anesthetics. Consequently, administering morphine will displace ether or chloroform to the sensory nerves and will cause anesthesia with less chloroform or ether than would be needed without the morphine.

In terms of the theory, the difference between an anesthetic like ether and a habit-forming drug like morphine must be that the ether is eliminated relatively rapidly, while the morphine or some of its reaction products are retained, keeping the system in the irritable stage. Experimentally, this has been found to be true. An obvious corollary is that a compound which would peptize the colloids of the sensory nerves and which had no deleterious effect on other parts of the organism should furnish relief to drug addicts. We are working on this problem now.

Nitrous oxide does not coagulate yeast and does not narcotize it. Nitrous oxide does not coagulate albumin and yet it is an anesthetic for man, being much used in dentistry. Nitrous oxide apparently interferes with the oxygen metabolism, giving rise to acid products which cause the flocculation. H. Wieland,[5] whose critical study on nitrous oxide led to the discovery of acetylene as an anesthetic, was the first to emphasize the necessity of separating nitrous oxide and acetylene from the narcotics of the ether and chloroform type. He points out that these gases produce an effect similar to mountain sickness. He is led to the view that there is an inhibition of oxidation especially in the brain.

The general conclusions to be drawn from our investigations are as follows:

1. Claude Bernard was right in postulating that reversible coagulation of the colloids of the sensory nerves produces or accompanies anesthesia.

2. Decreasing stability of the nerve colloids mean increasing irritability up to the point of incipient coagulation.

3. Claude Bernard's theory affords an apparently adequate explanation of the difference between a general and local anesthetic.

4. Claude Bernard's theory accounts for the effects observed when morphine is used together with chloroform or ether.

5. Claude Bernard's theory offers an apparently adequate explanation for the difference between a true anesthetic and a habit-forming drug.

6. Claude Bernard's theory accounts for the phenomena observed in the action of mercuric chloride on bacteria, of strychnine on frogs and dogs and of hydrocyanic acid on the eggs of citrus scale.

7. Decreasing stability of a suitable colloid or group of colloids—accompanied by increasing irritability up to a certain point and then reversible coagulation—will apparently account for the law, accepted as universal by biologists, physiologists, and bacteriologists that all depressor substances exert a stimulating effect when used at an appropriate dilution.

8. Claude Bernard's theory offers a clue to a scientific method of countering habit-forming drugs.

9. Nitrous oxide and acetylene are not true anesthetics in the sense that ether and chloroform are. These two gases cause reversible coagulations of the colloids of the sensory-nerves indirectly, and not directly as ether and chloroform do.

10. Claude Bernard's theory dealt with the colloids of sensory nerves. There is, therefore, nothing to prevent reversible coagulation of other materials besides proteins from being important in connection with anesthesia. It seems wise, however, to see how much can be explained solely by reversible coagulation of the proteins. We then have the behavior of the other colloids in reserve to be called into play when and as needed.

11. The Overton-Meyer theory accounts satisfactorily for many, though not for all, cases of cell permeability; but it is not true that all substances which enter the cells produce narcosis, and consequently the Overton-Meyer theory is a theory of permeability and not in any proper sense a theory of narcosis at all.

* This work is part of the programme now being carried on at Cornell University under a grant from the Heckscher Foundation for the Advancement of Research established by August Heckscher at Cornell University.

** NATIONAL RESEARCH FELLOW.

[1] Cf. *Leçons de l' anesthesie*, **1875.**

[2] Gurchot, *J. Phys. Chem.*, **30**, 83 (1926).

[3] Bancroft, *The Rice Institute Pamphlet*, **15**, 275 (1928).

[4] *Kolloid-Z.*, **11**, 209 (1912); *Troisième Congrès internationale de Neurologie et de Psychiatrie* (1913).

[5] *Archiv. exp. Path. Pharm.*, **92**, 96 (1922).

THE STRUCTURE OF THE CHLORITES

By Linus Pauling

Gates Chemical Laboratory, California Institute of Technology

Communicated July 9, 1930

By the application of the same methods of investigation as those used in the study of the micas,[1] I have determined the structure of the chlorites, a group of minerals morphologically similar but distinctly different in chemical composition. This structure is described in the following paragraphs.

Data obtained from oscillation and Laue photographs of specimens of penninite and clinochlore, of composition approximating the formula $Al_2Mg_5Si_3O_{16}(OH)_8$, lead to a monoclinic unit of structure with a = 5.2–5.3 Å., b = 9.2–9.3 Å., c = 14.3–14.4 Å., $\beta = 96°50'$, containing two molecules as written above. The lattice is apparently end-centered on (001), the space group being C_{2h}^3.

The close approximation of the values of a and b to those for the micas indicates that these crystals, too, are composed of tetrahedral and octahedral layers. The composition requires that in unit length along c there be two tetrahedral layers and two complete octahedral layers such as in brucite. Now we expect that as a rule the two faces of a constituent layer of a layer crystal will be equivalent, for if they were not equivalent there would be a tendency for the layer to curve, one face becoming concave and one convex, and this tendency would in general not be overcome by the relatively weak forces operative between adjacent layers. This symmetry is possessed by the layers of nearly all known layer structures—the micas, brittle micas, etc., brucite, hydrargillite, cadmium chloride, molybdenite and the A-modification of the sesquioxides. It is probable that kaolinite, $Al_2Si_2O_5(OH)_4$, is an exception. The very close approximation in dimensions of the octahedral layer in hydrargillite, $Al(OH)_3$, with b = 5.07 Å., a = 8.65 Å., and a complete layer of silicon tetrahedra, with a = 5.03 Å., b = 8.71 Å. (as in β-tridymite and β-cristobalite), shows that an unsymmetrical layer composed of these two, as in figure 1, would have only a very small tendency to curve, which could be overcome by the forces between layers. This structure for kaolinite (the different modifications of which would differ in the method of superposition of the layers) is supported by Hendrick's observation[2] that the value of d_{001} is about 7.1 Å., in agreement with prediction.

The non-existence of a magnesium analogue of kaolinite is accounted for by the large values of the fundamental translations in the brucite layer (with a = 5.40 Å.), which would cause the kaolinite-type layer to curve.

The chlorites are composed of mica layers and brucite layers alternating,

as shown in figure 2. Each of these layers is symmetrical. Aluminum tetrahedra expand the tetrahedral layers enough to allow them to be comfortably fitted on to a complete octahedral layer, as in biotite. The substitution of Al^{+3} for Si^{+4} gives the mica layers a resultant negative charge, as in the micas and brittle micas themselves. This is compensated by the alternating brucite layers, each of which through the sub-

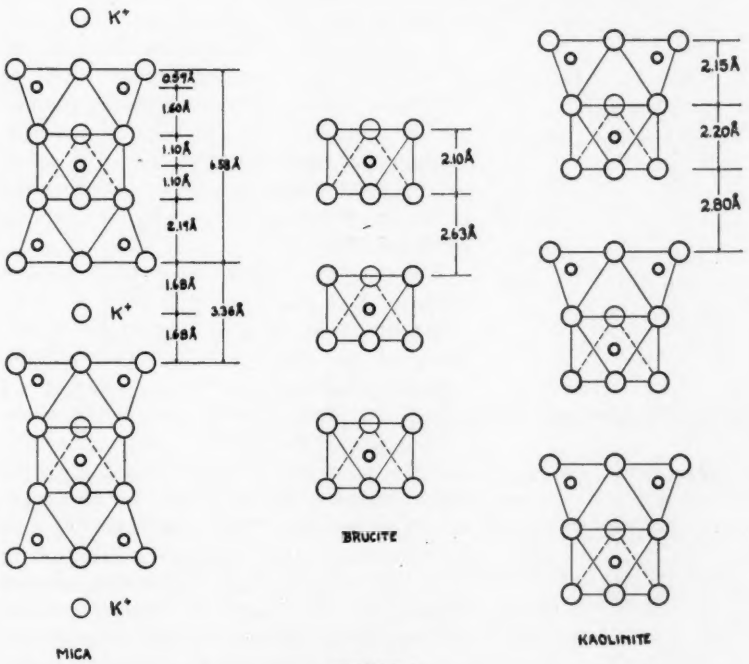


FIGURE 1

The structures of mica, brucite and kaolinite, showing the sequence of atom layers normal to the cleavage planes. Large circles represent O^- or OH^- (or K^+ when so marked), small ones Si^{+4} or Al^{+3} (at tetrahedron centers) and Mg^{++} or Al^{+3} (at octahedron centers). In each layer the atoms are arranged as described in Ref. 1.

stitution of Al^{+3} for Mg^{++} has a resultant positive charge, and which play the same rôle as K^+ and Ca^{++} in the micas and brittle micas.

Assuming that all the atoms in a layer have the same z coördinate, the succession of atoms normal to (001) is given by four parameters, with the predicted values $z_1 = 0.080$, $z_2 = 0.230$, $z_3 = 0.425$, and $u = 0.190$. The relative intensities of reflection in various orders from (001) are then determined by the amplitudes

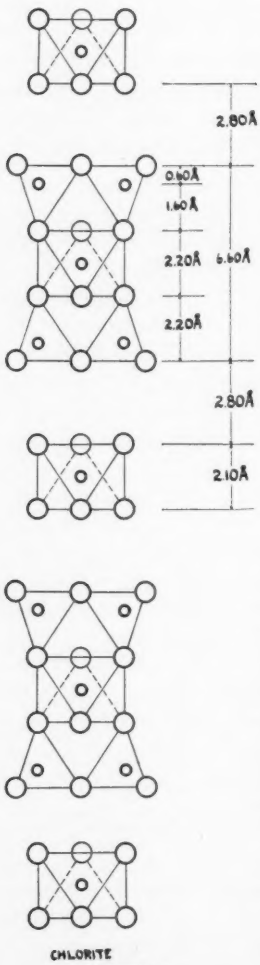


FIGURE 2

The sequence of atom layers normal to the cleavage plane of chlorite, showing mica layers of composition varying between $[Mg_3AlSi_3O_{10}\ (OH)_2]^-$ and $[Mg_3Al\ Al_2Si_2O_{10}\ (OH)_2]^-$, alternating with charged brucite layers, $[Mg_2Al(OH)_6]^+$.

$$A_{001} = 6A_{Mg,Al}\{1 + (-1)^l\} + 8A_{Si,Al}\cos 2\pi lu + 12A_O\{\cos 2\pi lz_1 + \cos 2\pi lz_2 + \cos 2\pi lz_3\}.$$

In table 1 are given values of $A^2/1600$ calculated in this way, together with the observed relative intensities of reflection in 26 orders from (001), obtained by the visual comparison of four photographs (made using MoK_α radiation) taken under identical conditions except for exposure times ranging from $8^1/_2$ min. to 1050 min. The comparison is shown graphically in figure 3. The agreement is excellent, verifying the suggested structure. The agreement is deleteriously affected by changing any of the parameters by as much as 0.01 from its predicted value.

The structure leads to the general chemical formula $X_mY_4O_{10}(OH)_8$, with $4 \leq m \leq 6$.

Actually m is equal to 6 for the great majority of the chlorites, which then contain complete octahedral layers only. With $m = 4$ the crystal would have the same composition as kaolinite; it is probable, however, that the kaolinite structure described above is the more stable, and that this lower limit for m is never closely approached. The constituent Y_4, ions with the coördination number 4, varies within the range between Si_3Al (as in penninite and clinochlore) and Si_2Al_2 (in amesite). The latter limit is imposed by the electrostatic valence rule; beyond it there would be some oxygen ions with the too low value $1^1/_2$ for Σs. With $m = 6$ the constituent X_m, ions with the coördination number 6, varies between Mg_5Al and Mg_4Al_2, with Fe^{++} and Mn^{++} sometimes partially replacing Mg^{++} and Fe^{+3} replacing Al^{+3}.

The existence of hexagonal layers of tetrahedra and octahedra in the chlorites accounts for their pronounced pseudohexagonal nature. The presence of layers not held together by anion-cation contacts explains their excellent cleavage on (001). The chlorites are not so soft

as kaolinite and talc because of the fact that the layers are charged alternatively positive and negative. On the other hand, the cleavage lamellae, while tough, are not elastic as in the micas, for the effect of the potassium ions in the micas in keying adjacent layers is here absent. The presence of

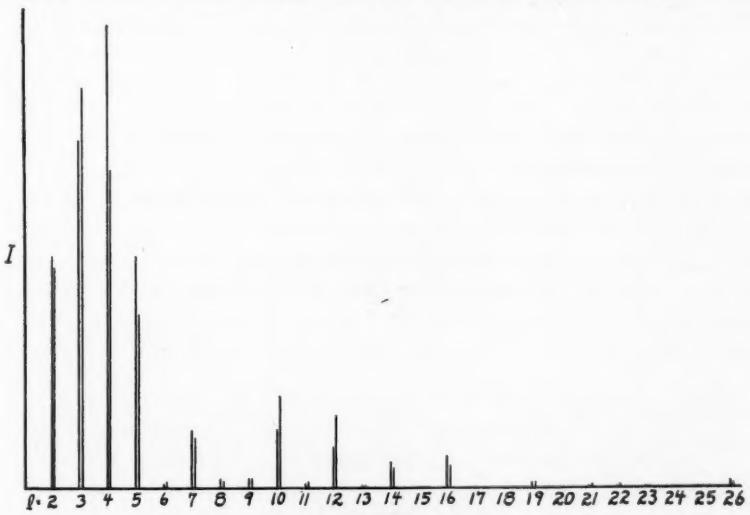


FIGURE 3

Comparison of the observed and calculated intensities of reflection of X-rays from (001) of penninite, in the 2nd. to 26th. order. Of each pair of vertical lines that on the left shows the observed intensity, that on the right the calculated intensity.

mica layers accounts for the possession by the chlorites of the same percussion and pressure figures as the micas, and also for the fact that chlorites and micas often form parallel growths and that the micas (particularly biotite) are easily altered into chlorites.

TABLE 1

OBSERVED AND CALCULATED INTENSITIES OF REFLECTION FROM (001) OF PENNINITE

Order of reflection	1	2	3	4	5	6	7	8	9	10	11	12	13
Observed intensity	4*	40	60	80	40	0.5	10	1.5	1.5	10	0.5	7	0.2
Calculated intensity	37	38	69	55	30	0.9	8.4	0.9	1.5	16	0.9	12.5	0.3
Order of reflection	14	15	16	17	18	19	20	21	22	23	24	25	26
Observed intensity	4.5	0.0	6	0.0	0.0	0.8	0.0	0.2	0.1	0.0	0.0	0.0	1
Calculated intensity	3.4	0.0	3.9	0.1	0.0	0.8	0.1	0.5	0.8	0.0	0.0	0.0	0.7

* The experimental arrangement used was such as to make the first order reflection abnormally weak.

It is evident that further crystals containing mica and brucite layers might exist. Representing a mica layer by M^-, a charged brucite layer by B^+ and a layer of potassium ions by K^+, the succession of structures

Mica:	$M^- K^+ M^- K^+ M^- K^+ M^- K^+$
Intermediate:	$M^- K^+ M^- B^+ M^- K^+ M^- B^+$
Chlorite:	$M^- B^+ M^- B^+ M^- B^+ M^- B^+$

indicates how intermediates between mica and chlorite might be formed. Giving M^- and B^+ the representative compositions $[Mg_3AlSi_3O_{10}(OH)_2]^-$ and $[Mg_2Al(OH)_6]^+$, the intermediate shown would have the composition $KMg_5Al_3Si_6O_{20}(OH)_{10}$, and would have an effective interplanar distance on (001) of 24.3 Å., just the sum of those for mica and chlorite. Chlorite-brucite intermediates would probably be unstable because of the juxtaposition of two positively charged brucite layers.

A detailed account of this investigation will be published in the *Zeitschrift für Kristallographie.*

[1] Linus Pauling, These PROCEEDINGS, **16**, 123–129 (1930).
[2] S. B. Hendricks, *Zeit. Kristallographie*, **71**, 269 (1929).

DISCUSSION OF VARIOUS TREATMENTS WHICH HAVE BEEN GIVEN TO THE NON-STATIC LINE ELEMENT FOR THE UNIVERSE

BY RICHARD C. TOLMAN

NORMAN BRIDGE LABORATORY OF PHYSICS, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated August 5, 1930

§1. *Introduction.*—In several previous articles,[1,2,3] I have shown the possibility of deriving a non-static cosmological line element to agree with the relatively uniform distribution of matter observationally found in the universe, and of using this line element to account both for the annihilation of matter taking place throughout the universe, and for the red-shift in the light from the extra-galactic nebulæ. After the publication of the first of these articles, I learned from a conversation with Professor H. P. Robertson that he had previously published[4] a different derivation of the same general form of line element; more recently Professor de Sitter and Sir Arthur Eddington have both been kind enough to call my attention to the previous use of a non-static line element by Lemaitre,[5] and still more recently I have discovered the early treatment of non-static line elements by Friedman.[6]

The purpose of the present article is to discuss certain features in this earlier work which overlap the treatment which I have given. Some remarks will also be made concerning an article by Eddington[7] which

appeared after the publication of my first article, and one by de Sitter[8] which appeared just after the publication of my second article, both of these having come to hand while the present note was in course of preparation.

§2. *The General Form of the Line Element.*—Employing the same symbolism which I have previously used, the non-static cosmological line element can be written in the general form

$$ds^2 = -\frac{e^{g(t)}}{[1 + r^2/4R^2]^2}(dr^2 + r^2d\theta^2 + r^2\sin^2\theta d\phi^2) + dt^2, \quad (1)$$

where R is a constant, and the dependence on the time arises from the function $g(t)$. There is an entire equivalence between this line element and those of Robertson, Lemaitre and Friedman. The translation of results between the different systems of symbols can be made with the help of the following equations in which the symbols on the left hand side correspond to equation (1) and those on the right hand side to the author indicated.

$$\left.\begin{aligned} R &= R \\ g(t) &= 2f(t) \\ \frac{r}{1 + r^2/4R^2} &= r \end{aligned}\right\} \begin{aligned} &\text{Robertson}^9 \\ &\\ &\text{Equation 10} \end{aligned} \quad (2)$$

$$\left.\begin{aligned} &Re^{g/2} = R \\ &\frac{1}{[1 + r^2/4R^2]^2}\left(\frac{dr^2}{R^2} + \frac{r^2}{R^2}d\theta^2 + \frac{r^2}{R^2}\sin^2\theta d\phi^2\right) = d\sigma^2 \end{aligned}\right\} \begin{aligned} &\text{Lemaitre} \\ &\\ &\text{Equation 1} \end{aligned} \quad (3)$$

$$\left.\begin{aligned} Re^{g/2} &= R(x_4)/c \\ \frac{r/R}{1 + r^2/4R^2} &= \sin x_1 \end{aligned}\right\} \begin{aligned} &\text{Friedman} \\ &\\ &\text{Equation } D_3 \end{aligned} \quad (4)$$

The derivation which I gave to this line element was based on five assumptions. Four[10] of these assumptions could be regarded as direct consequences of the approximately uniform distribution of material in the universe observationally found to those depths to which the Mount Wilson instruments have been able to penetrate, and the fifth[11] was the stability requirement that nebulæ which are stationary *in the system of coordinates used* should not be subject to acceleration but should remain permanently stationary.

The derivation of Robertson was based on two assumptions, first that space-time should be separable into space and time, and second that it

should be spatially homogeneous and isotropic. The latter of these assumptions is closely related to my four assumptions that flow from the observational approximately uniform distribution of material in the universe, and the former has the same effect on the form of the line element as my stability requirement. There is thus a close relation between the starting points of Robertson and myself. I should wish to emphasize, however, that it would seem to me premature if we should assert at present that the whole of space-time is necessarily even approximately spatially homogeneous. I prefer to regard the line element (1) as an approximate expression in agreement with the observational uniform distribution of matter out to the range of the 100-inch telescope, but one which could be reasonably modified to take account of different conditions if they should be found at greater distances.

Lemaitre presents no derivation of the line element used, and the derivation given by Friedman appears less satisfying than those given by Robertson and myself, especially as he introduces at the very start an assumption (*C*) which would rule out the presence of radiation in the universe.

§3. *Treatment of the Red Shift by Lemaitre.*—The non-static line element under discussion was not applied to observational data by Robertson or Friedman, but was used by Lemaitre to account for the red shift in the light from extra-galactic nebulæ. In doing this he introduced two assumptions—first that the proper pressure in the universe should be permanently zero, and second that the cosmological constant occurring in Einstein's equations should have a particular value that would make the minimum proper volume of the universe occur at the time minus infinity. In the present section we shall discuss the results of these assumptions, and then in the following section compare the treatment of Lemaitre with treatments which I have tried.

a. Effect of Zero Pressure on Presence of Radiation.—The assumption of zero proper pressure at once excludes the presence of radiation in the universe, for which the pressure is one-third the energy density. Hence this assumption presumably does not agree precisely with the facts, since the universe appears to have had a long past history in which radiation has been emitted from the heavenly bodies. Moreover, it is interesting to note in this connection that the temperature of intergalactic space would only have to be of the order of one degree absolute to give a density of radiation equal to Hubble's value for the averaged out density of matter in the universe, 1.5×10^{-31} gm./cc. Nevertheless, we have no easy means of estimating the actual pressure, and can make no precise quantitative judgment as to the degree of correctness of Lemaitre's assumption on this point.

b. Effect of Zero Pressure on the Annihilation of Matter.—The as-

sumption that the pressure in the universe is zero and hence that radiation is permanently excluded would also rule out the annihilation of matter, that is, its transformation into radiation. This can be seen from my general equation[12] for the fractional rate of annihilation of matter

$$-\frac{1}{M}\frac{dM}{dt} = 6\frac{p_0}{\rho_0}\dot{g} + \frac{3}{\rho_0}\frac{dp_0}{dt}, \qquad (5)$$

where p_0 is the proper pressure in the universe, ρ_0 the proper density of matter, and $\dot{g} = dg/dt$. With the pressure p_0 permanently zero, both terms on the right hand side of the equation (5) are zero and the amount of matter in the universe would be a constant.

This result also does not appear to be in precise agreement with the facts, which seem to indicate a continuous annihilation of matter in order to maintain the radiation from the stars. In addition, we have some indications that an annihilation of matter in intergalactic dust may be associated with the production of the cosmic rays.

Moreover, as I have previously pointed out the annihilation of matter has the effect of making the line element of the universe necessarily non-static and hence necessarily leading to a shift in the wave length of light from distant objects, and, as I also showed, a red rather than a violet shift if we make use of the additional assumption that the proper motions of the nebulæ are tending to decrease rather than increase with the time.[13]

There are thus certain further arguments against Lemaitre's assumption of zero pressure since it would lead to a zero rate of annihilation of matter. Nevertheless, since we have no satisfactory theory of the enormous variation with stellar mass in the rates of annihilation actually observed in the stars, and have no satisfactory knowledge as to the possible rate of annihilation of matter in intergalactic dust, it is not possible to decide exactly what the actual figure for the average rate of annihilation should be, or to make any precise quantitative judgment of Lemaitre's hypothesis from this point of view.

c. Effect of Zero Pressure on the Form of the Line Element.—The assumption of zero pressure prescribes a definite dependence of the line element on the time. In accordance with the principles of relativistic mechanics the proper pressure corresponding to the line element (1) is given by the equation[14]

$$8\pi p_0 = -\frac{1}{R^2}e^{-g} - \ddot{g} - \frac{3}{4}\dot{g}^2 + \Lambda, \qquad (6)$$

where Λ is the cosmological constant and the dots indicate differentiation with respect to the time. Setting this expression equal to zero we can obtain as a first integral

$$\dot{g} = \sqrt{\frac{4\Lambda}{3} - \frac{4}{R^2}\, e^{-g} + Ae^{\frac{-3g}{2}}} \tag{7}$$

where A is a constant of integration, and can then write the complete integral in the general form

$$t = \int \frac{dg}{\sqrt{\frac{4\Lambda}{3} - \frac{4}{R^2}\, e^{-g} + Ae^{-\frac{3g}{2}}}}. \tag{8}$$

d. Effect of Lemaitre's Second Assumption on the Form of the Line Element.—If we now introduce Lemaitre's second assumption, as to the magnitude of Λ, the indicated integral in equation (8) can be explicitly evaluated. This second assumption is given by the equation

$$A = \frac{8}{3R^3\, \Lambda^{1/2}}. \tag{9}$$

Substituting in (8), and for simplicity of expression, writing in agreement with the notation of Lemaitre

$$\Lambda = \frac{1}{R_0^2} \tag{10}$$

we can obtain[15]

$$t = R_0 \sqrt{3} \log \frac{\sqrt{Re^{g/2} + 2R_0} + \sqrt{Re^{g/2}}}{\sqrt{Re^{g/2} + 2R_0} - \sqrt{Re^{g/2}}} + R_0 \log \frac{\sqrt{3Re^{g/2}} - \sqrt{Re^{g/2} + 2R_0}}{\sqrt{3Re^{g/2}} + \sqrt{Re^{g/2} + 2R_0}} + \text{const.} \tag{11}$$

as an explicit expression for the dependence of g on t.

In accordance with this result the value of $Re^{g/2}$ could be regarded as increasing continuously from a minimum value R_0 with which it started at a time infinitely remote in the past, thus providing the long period for past stellar evolution desired by Lemaitre.[16]

e. Effect of Form of Line Element on Relation between Red Shift and Distance.—As shown in my previous work a linear relation between red shift and distance is to be obtained by taking a linear dependence of g on t. Hence the question naturally arises whether such a complicated relation between g and t as that given by Equation (11) would give a dependence of red shift on distance nearly enough linear to agree with the actual observations of Hubble and Humason on the red shift in the light from the nebulæ. This, however, proves to be the case.

As an exact expression for the dependence of fractional red-shift on the coordinate r we have[17]

$$\frac{d}{dr}\left(\frac{\delta\lambda}{\lambda}\right) = \frac{1}{1 + r^2/4R^2}\,\frac{\dot{g}}{2}, \tag{12}$$

where $\dot{g}$ is the value of dg/dt at the time the light was emitted. With sufficient approximation for our present purposes this can be written as

$$\frac{d}{dL}\left(\frac{\delta\lambda}{\lambda}\right) = \frac{\dot{g}}{2}, \tag{13}$$

where L is to be taken as agreeing with the observational distances of Hubble, and with g linear in t this gives us the desired linear dependence of red shift on distance.

As the observational value for the relation between red shift and distance Lemaitre (Equation 24) takes

$$\frac{d}{dL}\left(\frac{\delta\lambda}{\lambda}\right) = \frac{\dot{g}}{2} = 6.4 \times 10^{-10}\ (\text{years})^{-1}. \tag{14}$$

To estimate the deviation from linearity involved in the treatment of Lemaitre, we may return to the expression for $\dot{g}$ given by equation (7) and by differentiating with respect to the time and substituting equations (9) and (10) obtain

$$\frac{\ddot{g}}{2} = \frac{1}{R^2 e^g} - \frac{R_0}{R^3 e^{3g/2}}. \tag{15}$$

With the values of $Re^{g/2}$ and R_0 of Lemaitre this gives us

$$\frac{\ddot{g}}{2} \approx 3 \times 10^{-21}\ (\text{years})^{-2}. \tag{16}$$

On the other hand the most remote nebula whose red shift has so far been determined is at a distance of approximately 10^8 light years, so that the change in $\dot{g}/2$ or in the slope of $\delta\lambda/\lambda$ as a function of L in going out to the most distant nebula observed would be of the order

$$\Delta\left[\frac{d}{dL}\left(\frac{\delta\lambda}{\lambda}\right)\right] \approx 3 \times 10^{-13}\ (\text{years})^{-1}, \tag{17}$$

which for the present is negligible compared with the observational value 6.4×10^{-10} (years)$^{-1}$. This is in agreement with the fact specially pointed out in my second article that it is possible to assign a small positive value to $\ddot{g}$ which will be large enough to reduce the rate of annihilation of matter to any desired extent below that calculated from a linear depend-

ence of g on t, and yet small enough not to affect seriously the linear dependence of red shift on distance.

§4. *Comparison of Treatments by Lemaitre and the Author.*—It will now be interesting to compare the treatment of the non-static line element by Lemaitre with the treatments given in my first and second papers and certain other possibilities. This can be done with the help of the following table which presents some of the significant features of these different treatments.

CASE	PRESSURE	FORM OF g	ANNIHILATION	RED SHIFT
L	$p_0 = 0$	Equation 11	$-\frac{1}{M}\frac{dM}{dt} = 0$	$\frac{\delta\lambda}{\lambda} \backsimeq kL$
T_1	$p_0 \ll \rho_0$	$g = 2kt$	$-\frac{1}{M}\frac{dM}{dt} = 3k$	$\frac{\delta\lambda}{\lambda} = kL$
T_2	$p_0 \ll \rho_0$	$g = 2(kt + lt^2 + mt^3 \ldots)$	$-\frac{1}{M}\frac{dM}{dt} < 3k$	$\frac{\delta\lambda}{\lambda} \backsimeq kL$
T_3	$p_0 \gg \rho_0$	$g = 2kt$	$-\frac{1}{M}\frac{dM}{dt} \gg 3k$	$\frac{\delta\lambda}{\lambda} = kL$
T_4	$p_0 \ll \rho_0$	$e^g = (1 + kt)^2$	$-\frac{1}{M}\frac{dM}{dt} = 3k$	$\frac{\delta\lambda}{\lambda} \backsimeq kL$
T_5	$p_0 = \text{const.}$	Equation 11	$-\frac{1}{M}\frac{dM}{dt} = \frac{12p_0}{\rho_0}k$	$\frac{\delta\lambda}{\lambda} \backsimeq kL$

The first column in this table gives a symbol by which the case can be designated. The second and third columns give the assumptions which are made as to the pressure in the universe and the form of dependence of the line element on the time, some assumption on these points being necessary in order to permit any application at all. The last two columns give the resulting value for the calculated rate of annihilation of matter, and the resulting dependence of the red shift on distance.

Case L is that of Lemaitre. It has the disadvantage of the very complicated dependence of g on t, given by equation (11), which still remains complicated in the symbolism of Lemaitre. It also suffers from the assumptions of absolutely zero pressure in the universe and absolutely zero rate of annihilation of matter as discussed above. Nevertheless our actual knowledge as to pressure and rate of annihilation is insufficient to rule it out as possibly being a good approximation.

Case T_1 is the one presented in my first paper, and the first of those discussed in my second paper. It has the advantage of the simplest possible dependence of g on t which leads immediately to a linear dependence of red shift on distance. It also has the possible advantage that it makes the logarithmic rate of annihilation of matter a constant

so long as $p_0 \ll \rho_0$. Such a logarithmic rate of decay would seem theoretically reasonable except for the actual complications found in the rate of decay of matter in stars of different masses. The case in question gives a rate of annihilation of matter as high as that observationally found in the most massive stars, far higher than that in a typical star such as the sun. If, however, it should be found that the cosmic rays arise from an exceedingly high rate of annihilation of intergalactic dust, this case might be an approximately correct description of the present state of the universe.

Case T_2 is the second of those discussed in my second paper, and gives perhaps the most satisfactory treatment of our present knowledge of the state of the universe now possible. The pressure is still taken small compared with the density of matter, but the dependence of g on t is modified by the possibility of terms in t of higher order than the first. In particular it was shown in my second paper that it was possible to assign a small positive value to the coefficient l in the expression $g = 2(kt + lt^2)$, which would be big enough to reduce the calculated value for the present rate of annihilation of matter as far as desired, and yet at the same time small enough not to affect appreciably the linear relation between red shift and distance out to a depth of 10^8 light years, which is as far as the observations of Hubble and Humason have yet been pushed.

Case T_3, which was also considered in my second paper, is interesting in illustrating the effect of assuming that the density of radiation in the universe is large compared with the averaged out density of matter. It gives an exceedingly high rate of annihilation of matter because of the great effect (with $p_0 \gg \rho_0$) of the first term on the right-hand side of the equation previously cited

$$-\frac{1}{M}\frac{dM}{dt} = \frac{6p_0}{\rho_0}\cdot g + \frac{3}{\rho_0}\frac{dp_0}{dt}. \qquad (18)$$

In general, the assumption of a pressure in the universe large compared with the density of matter would seem difficult.

Case T_4, which has not previously been published, is interesting in providing another very simple dependence on the time which gives approximately the same results as Case T_1.

Case T_5, which has also not been previously published, is of interest in taking the pressure in the universe as constant. This might have an approximate validity since there are some processes (emission of radiation) which tend to increase the pressure and others (expansion of the universe) which tend to decrease it. With p_0 constant we can take the relation between g and t as that presented by equation (11), given above for the Lemaitre theory, with the change that corresponding to equations (9) and (10) we should have instead

$$A = \frac{8}{3R^3\sqrt{\Lambda - 8\pi p_0}} \quad \text{and} \quad \Lambda - 8\pi p_0 = \frac{1}{R_0^2} \tag{19}$$

§5. *Remarks Concerning the Articles of Eddington and de Sitter.*—After the publication of my first paper, an article appeared by Eddington (loc. cit.) making use of the non-static line element to discuss the stability of the Einstein universe. In this article he has incidentally made two criticisms of my first paper with which I cannot entirely agree.

Referring perhaps to the title of my paper, "The Effect of the Annihilation of Matter on the Wave-Length of Light from the Nebulæ," Eddington says that the conversion of matter into radiation cannot be regarded as an explanation of the recession of the nebulæ, since it tends to retard the expansion of the universe.

I am quite willing to agree that the title of my paper was in some ways unfortunately chosen, since in the case of a number of mathematically interrelated phenomena, it is dangerous to label one of them cause and the rest effects, and since as shown by Lemaitre the non-static line element can also be applied to an expanding universe without introducing any annihilation of matter.

I should wish to reëmphasize, however, that the annihilation of matter necessarily leads to a non-static line element, and to a universe in which there is a red rather than a violet shift if we are willing to exclude the possibility that the proper motions of the nebulæ tend to increase with the time. I should also wish to emphasize that the annihilation of matter must certainly be accompanied either by an expansion of the universe or by a rise in pressure. This is very clearly shown by the equation already cited

$$-\frac{1}{M}\frac{dM}{dt} = \frac{6p_0}{\rho_0}\frac{\dot{g}}{g} + \frac{3}{\rho_0}\frac{dp_0}{dt}, \tag{20}$$

which makes it evident that a positive value for the fractional rate of annihilation $(-1/M)(dM/dt)$ must be accompanied either by an expanding universe, i.e., g positive, or by a rising pressure, i.e., dp_0/dt positive.

The second criticism which Eddington made was based on the fact that in my first paper I assumed a simple linear formula $g = 2kt$ to express the dependence of the line element on the time. It was definitely stated in the article, however, that this was to be regarded as a first approximation, and since then I have discussed the introduction of higher terms in my second paper. It may be remarked, moreover, that as far as the motions of the nebulæ are concerned the simple linear formula appears to be a good approximation out to 10^8 light years. It cannot be stated, however, that it is necessarily a good approximation for the calculation of rates of annihilation.

Shortly after the publication of my second paper, an interesting treat-

ment of the non-static line element for the universe was published by de Sitter (loc. cit.). In this treatment he first considers the forms of the line element that would correspond to a universe with no annihilation of matter, but not necessarily with zero pressure. The different forms considered correspond to different choices of the parameters. The forms are very complicated, that of Lemaitre as given by equation (11) being a specially simple case. De Sitter then considers the possibility of allowing for the annihilation of matter by permitting one of the parameters in his previous equations to vary with the time in such a way as to give a rate of annihilation of the order of that found in a dwarf star of a somewhat later type than the sun or even less. In this connection he remarks in a footnote on the high rate of annihilation given by the treatment in my first paper, but by this time my second paper had already been published showing the possibility of obtaining as low a rate of annihilation as desired by introducing terms of higher order than the first in the expression for the dependence of g on t.

From a practical point of view the treatment of de Sitter suffers from the great complexity in the forms of line element obtained, and from a theoretical point of view in the fact that they correspond to a universe in which no annihilation of matter takes place, and the allowance for annihilation is made as a later correction without inclusion in the explicit form of the line element. Furthermore, although his forms of line element apparently give the whole history of the expansion of the universe, he himself emphasizes that he does not wish to lay stress on the implications for the distant past or future, and determines the parameters for his different forms from considerations of the present density and rate of expansion of the universe. Hence in some ways, I prefer the treatment which had already been published in my second article, Case T_2 of the table, in which we assume a simple expansion in powers of t, which only attempts to represent the dependence of the line element on the time in the neighborhood of the present, and in which as many parameters as possible are to be determined from our present knowledge of density and pressure, rate of expansion, *and* rate of annihilation.

NOTE: Another article by de Sitter (*Proc. Nat. Acad.*, **16**, 474, 1930) has arrived just as this paper was to be sent to the Editor. It contains, however, no further material which needs comment from the present point of view.

§6. *Conclusion.*—In conclusion it would seem as if our present knowledge of the universe could best be treated by the methods of my second article in accordance with the Case T_2 of the table, by taking the cosmological line element in the form

$$ds^2 = -\frac{e^{2(kt+ll^3)}}{[1+r^2/4R^2]^2}(dx^2+dy^2+dz^2)+dt^2, \qquad (21)$$

where $r = \sqrt{x^2 + y^2 + z^2}$ and R, k and l are constant parameters.

As I have shown this line element corresponds to a uniform distribution of material in the universe, and nebulæ stationary in the coördinates x, y, z will remain so. Furthermore, in the case of an expanding universe, with $(kt + lt^2)$ positive and the accompanying red shift from distant objects, the proper motions of the nebulæ would tend to decrease with the time.

At the time $t = 0$ (i.e., the present) the density of matter ρ_0 and the pressure p_0 would be related to the parameters by[18]

$$4\pi(\rho_0 + 4p_0) = \frac{1}{R^2} - 2l. \tag{22}$$

The dependence of the red shift on distance would be approximately linear provided l is sufficiently small[19]

$$\frac{\delta\lambda}{\lambda} \approx kL. \tag{23}$$

And the rate of annihilation of matter would be given by[20]

$$-\frac{1}{M}\frac{dM}{dt} = 12\frac{p_0}{\rho_0}k + \left(1 + \frac{4p_0}{\rho_0}\right)\frac{\frac{3}{R_2} - 18l}{\frac{1}{R^2} - 2l}k. \tag{24}$$

We thus have three equations to determine the three parameters R, k and l in terms of physically measurable quantities. Taking the pressure p_0 small compared with the density of matter ρ_0, and using Hubble's estimate for the average density of matter in the universe we have

$$\frac{1}{R^2} - 2l \approx 10^{-22}\ (\text{years})^{-2}; \tag{25}$$

from the red shift we have

$$k = 5.1 \times 10^{-10}\ (\text{years})^{-1}, \tag{26}$$

which at present is the most accurately determined of the parameters, and finally with p_0 small compared with ρ_0, l can be chosen to give a rate of annihilation as small as that for example from the late dwarf stars, without appreciably affecting the linear relation between red shift and distance. Any actual choice for l, however, should perhaps wait until we have more knowledge as to the rate of annihilation of intergalactic dust. At the present moment, the use of any more complicated expression for the dependence of the line element on the time, or the introduction of higher temporal terms than lt^2 would perhaps be premature.

Further progress in the use of the non-static line element for describing the behavior of matter in the universe would seem possible both along theoretical and observational lines.

On the theoretical side we should be greatly helped by some acceptable interpretation and determination of the cosmological constant Λ, and this might result from a successful unified field theory, or from the proper fusion of general relativity and quantum mechanics. A satisfactory theory which would account for the enormous range in the rates of annihilation of matter found in stars of different mass would also be extremely helpful.

On the observational side more information as to the cause and character of the cosmic rays would of course be desirable. Primarily, however, our observational interests must lie in the furtherance of the remarkable determinations of the relation between red shift and distance which are being made by Hubble and Humason. It is perhaps too much to hope that they could obtain an observational accuracy which would detect the small deviations from linearity of red shift with distance which would be sufficient to account for a low rate of annihilation of matter. Even with the two-hundred inch telescope under construction for this Institute such an accuracy would perhaps still be impossible. However, it is a matter of the gravest interest to make sure that no large deviation from linearity occurs. And in addition it is not premature to hope that sufficient accuracy can be obtained in the measurement of the angular extensions and the luminosities of the nebulæ to make it possible to test the interrelation between these two methods of estimating distance, which was shown in my third paper to be a necessary consequence of the non-static line element. This will furnish a significant although not very extensive test of the general idea that the red shift is to be explained by a non-static line element.

[1] Tolman, *Proc. Nat. Acad. Sci.*, **16**, 320 (April 15, 1930).

[2] Tolman, *Ibid.*, **16**, 409 (June 15, 1930).

[3] Tolman, *Ibid.*, **16**, 511 (July 15, 1930).

[4] Robertson, *Ibid.*, **15**, 822 (1929).

[5] Lemaitre, *Ann. Société Sci. Bruxelles*, **47**, Series A, 49 (1927).

[6] Friedman, *Zeit. Phys.*, **10**, 377 (1922).

[7] Eddington, *Monthly Notices R. A. S.*, **90**, 668 (May, 1930).

[8] De Sitter, *Bull. Astron. Inst. of the Netherlands*, **5**, 211 (June 24, 1930).

[9] Robertson's r is the same as my $\bar{r}$ in the form of the line element given by equation (5), Reference 3.

[10] Reference 2, Assumptions *a*, *b*, *d*, *e*, in §2.

[11] Reference 2, Assumption *c* in §2.

[12] Reference 2, Equation (14).

[13] Reference 2, §4, 5.

[14] Reference 1, Equations 34; Reference 2, Equations 2. Note the omission of a dot in the latter place.

[15] This corresponds to Lemaitre's equation (30).

[16] The article of Friedman is interesting in considering a number of possibilities as to the past history of the universe, including that of a periodic expansion and contraction, all, however, with the restriction $p_0 = 0$. De Sitter also considers a number of such possibilities, all, however, with the restriction $dM/dt = 0$.

[17] Reference 2, Equation 23. This equation was derived setting the constant $g_2 = 0$. This is also desired for our present purposes, however, to make our equation (13) agree with Lemaitre's equation (24).

[18] Reference 2, § 8 (*e*). Note that this should read $4\pi(\rho_0 + 4p_0)$ or $4(\rho_{00} + p_0)$ instead of $4\pi(\rho_0 + p_0)$.

[19] Reference 2, §8 (*g*).

[20] Reference 2, Equation 28.

ON THE SUMMABILITY OF FOURIER SERIES. THIRD NOTE

BY EINAR HILLE AND J. D. TAMARKIN

PRINCETON UNIVERSITY AND BROWN UNIVERSITY

Communicated August 8, 1930

1. In the present note we continue the discussion of the summability of Fourier series by the methods $[H, q(u)]$ of Hurwitz-Silverman-Hausdorff.[1] We give necessary and sufficient conditions for effectiveness, and show that the problem is closely connected with the theory of Fourier transforms.

We recall that the function $q(u)$ associated with the method $[H, q(u)]$ is supposed to satisfy Hausdorff's conditions: $q(u)$ is of bounded variation in (0, 1) and continuous at $u = 0$, $q(0) = 0$, $q(1) = 1$. Let $f(x)$ be any integrable function of period 2π. The n^{th} (H, q)-mean of the Fourier series of $f(x)$ is

$$H_n[f(x), q] = \int_{-\pi}^{+\pi} f(t + x)H_n(t)\, dt, \tag{1}$$

where

$$H_n(t) = \frac{1}{\pi}\Re\left\{(1 - e^{it})^{-1}\int_0^1 (1 - u + ue^{it})^n dq(u).\right\} \tag{2}$$

We put

$$\varphi(t) = f(x + t) + f(x - t) - 2S, \tag{3}$$

$$\psi(t) = \int_0^t \varphi(\tau)d\tau, \quad \Phi(t) = \int_0^t |\varphi(\tau)|d\tau. \tag{4}$$

We shall use the following terminology. A point x will be called *regular* or *pseudo-regular* with respect to $f(x)$ according as there exists an $S = S_x$,

such that $\lim_{t \to 0} \varphi(t) = 0$ or $\Phi'(0) = 0$. It is well known that almost every point, and, in particular, every regular point is pseudo-regular. The method (H, q) will be said to be Fejér or (F) *effective*, resp. Lebesgue or (L) *effective* with respect to Fourier series if $\lim_{n \to \infty} H_n[f(x), q] = S_x$ for every integrable function $f(x)$ at every regular point, resp., at every pseudo-regular point of the function.

We introduce the following auxiliary functions. We denote the total variation of $q(u)$ in the interval $(0, u)$ by $Q(u)$. Further

$$\gamma(u) = \begin{cases} 0 & , \ 1 < u, \\ q(u) - u & , \ 0 \leq u \leq 1, \\ \gamma(-u) & , \ u < 0; \end{cases} \tag{5}$$

$$\Gamma(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-i\xi u} \gamma(u) du = \sqrt{\frac{2}{\pi}} \int_0^1 \cos \xi u \ \gamma(u) du; \tag{6}$$

$$\rho(\xi) = \int_0^1 \sin \xi u \, dq(u). \tag{7}$$

Finally, a function $g(u)$ is said to belong to the class L^p, $1 \leq p \leq 2$, (abbreviated $g(u) \subset L^p$) if it is integrable over any finite interval, and the integral

$$\int_{-\infty}^{+\infty} |g(u)|^p \, du$$

exists. We denote the Fourier transform of $g(u)$, which is known to exist, by $G(\xi)$. Thus, $\Gamma(\xi)$ is the Fourier transform of $\gamma(u)$.

2. After these preliminaries we proceed to the subject matter of this note.

It is easily verified that the condition

$$\int_0^\pi |H_n(t)| dt \leq M \ (n = 1, 2, \ldots) \tag{8}$$

is necessary and sufficient in order that (H, q) be (F) effective. The interval $(0, n^{-1})$ may be disregarded in discussing the boundedness of the integral. On the interval $(n^{-1}, n^{-1/2})$ it can be shown that

$$H_n(t) = \frac{1}{t} \rho(nt) + O(nt). \tag{9}$$

Hence

I. *A necessary condition for* (H, q) *to be* (F) *effective is that*

$$\Gamma(\xi)/\xi \subset L, \tag{10}$$

which is the case if and only if

$$\Gamma(\xi) \subset L. \tag{11}$$

Using the inequality

$$|1 - u + u e^{it}| < \exp\left[-\frac{2}{\pi^2} u(1-u)t^2\right],$$

we can show that the contribution of the interval $(n^{-1/2}, \pi)$ to the value of the integral in (8) remains bounded if

$$\lim_{\epsilon \to 0} \int_{\epsilon}^{1-\epsilon} \log \frac{1}{u(1-u)}\, dQ(u) < \infty, \tag{12}$$

but not necessarily only if this condition is satisfied.

II. *A sufficient condition for* (H, q) *to be* (F) *effective is that* $q(u)$ *satisfies* (10) *and* (12).

It should be noted that (10) implies that $q(u)$ is continuous in (0, 1) or, at least, equivalent to a continuous function. This may be proved in a variety of ways. The simplest and most direct proof is based on the reciprocal formula

$$\gamma(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{iu\xi} \Gamma(\xi) d\xi \tag{13}$$

which shows that $\gamma(u)$ is continuous if $\Gamma(\xi) \subset L$.

3. For (L) effectiveness we have to make additional assumptions. It is readily shown that:

III. *A necessary condition for* (H, q) *to be* (L) *effective is that*

$$\lim_{n \to \infty} n \int_{n^{-1}}^{n^{-1/2}} \varphi(t)\Gamma(nt)dt = 0 \tag{14}$$

at every pseudo-regular point. This condition together with (12) *is also sufficient.*

(14) may be replaced by the equivalent condition

$$\lim_{n \to \infty} n^2 \int_{n^{-1}}^{n^{-1/2}} \psi(t)\Gamma'(nt)dt = 0 \tag{15}$$

at every pseudo-regular point. (15) will be satisfied if, e.g.,

$$\xi\Gamma'(\xi) \subset L. \tag{16}$$

IV. *A sufficient condition for* (H, q) *to be* (L) *effective is* (16) *together with* (12).

It should be noticed that (11) and (16) together imply

$$\lim_{\xi \longrightarrow \pm \infty} \xi \Gamma(\xi) = 0. \tag{17}$$

It will be shown below that there exist (F) effective methods (H, q) for which (17), and, *a fortiori*, (16) do not hold.

4. For a further discussion of the problem it is clearly necessary to know when the Fourier transform $G(\xi)$ of a function $g(u) \subset L^p$ has the property of $\subset L$. A preliminary inspection of the literature has failed to yield any direct information about this question with the exception of a theorem due to Titchmarsh[2] which we may formulate as follows:

1. $G(\xi) \subset L$ if $g(u) \subset L^p$, $1 < p \leq 2$, and

$$\int_{-\infty}^{+\infty} |g(u + h) - g(u - h)|^p du = O(|h|^{\alpha p}), \; 0 < \alpha \leqq 1, \tag{18}$$

when $h \longrightarrow 0$, provided $\alpha p > 1$.

On the other hand, it is obvious that most of the known theorems about absolutely convergent Fourier series should have analogues for Fourier transforms. Thus, working by analogy, we obtain Theorems **2**–**5** below. Of these **4,** which is the analogue of Bernstein's theorem,[3] is an immediate consequence of **1; 5** is the analogue of Zygmund's theorem[4] and belongs to the same order of ideas as **1.** In addition we state theorem **6** which is closely related to the main theorem of our Second Note. Its analogue for Fourier series would seem to be new.

2. *A necessary and sufficient condition for* $G(\xi)$ *to* $\subset L$ *when* $g(u) \subset L^2$ *is that*

$$g(u) = \int_{-\infty}^{+\infty} g_1(t) g_2(u - t) dt, \tag{19}$$

where $g_1(t)$ *and* $g_2(t) \subset L^2$.[5]

Sufficient conditions for $G(\xi)$ to $\subset L$ when $g(u) \subset L^2$ (or L^p, $1 \leqq p \leqq 2$, in cases **3** and **6**) are:

3. $g(u)$ *and its conjugate* $\bar{g}(u)$ *are of bounded variation in* $(-\infty, +\infty)$. *Here*

$$\bar{g}(u) = \lim_{\epsilon \longrightarrow 0} \int_{\epsilon}^{\infty} [g(u + t) - g(u - t)] \frac{dt}{t}.^{6} \tag{20}$$

4. $g(u)$ *satisfies a Lipschitz condition of order* $\alpha > 1/2$ *uniformly in* $(-\infty, +\infty)$.

5. $g(u)$ *is of bounded variation and continuous in* $(-\infty, +\infty)$, *its modulus of continuity* $\omega(h)$ *being such that*

$$\int_0^1 [\omega(h)]^{1/2} \frac{dh}{h} < \infty \tag{21}$$

6. *$g(u)$ is of bounded variation in $(-\infty, +\infty)$ and*

$$\int_0^1 \Omega(h) \frac{dh}{h} < \infty \tag{22}$$

where

$$\Omega(h) = \int_{-\infty}^{+\infty} |d[g(u+h) - g(u-h)]|. \tag{23}$$

5. With the aid of these results we can construct methods (H, q) which are at least (F) effective. We shall merely give a couple of examples utilizing Theorems **2** and **5.**

Let $\gamma_1(t)$ and $\gamma_2(t)$ be even functions $\subset L^2$, $\gamma_\nu(t) = 0$ if $|t| > 1/2$, $\nu = 1, 2$. Further

$$\int_0^{1/2} \gamma_1(t)\ \gamma_2(t)dt = 0.$$

Putting

$$\gamma(u) = \int_{-\infty}^{+\infty} \gamma_1(t)\gamma_2(u-t)dt,\ q(u) = u + \gamma(u), \tag{24}$$

and assuming in addition that $\gamma_2(t)$ is of bounded variation, we obtain a method (H, q) which is (F) effective.

The following theorem is a consequence of Theorems II and **5:**

V. *A sufficient condition for (H, q) to be (F) effective is that $Q(u)$ is continuous and its modulus of continuity satisfies condition* (21).

An interesting example of a non-absolutely continuous function satisfying the conditions of Theorem V but not condition (16), is presented by the so-called Cantor-Lebesgue function.[7] We have here an (F) effective method (H, q) which may conceivably not be (L) effective, but we are unable at present to give a definite answer to this question.

[1] These PROCEEDINGS, **12**, 915–918 (1928), and **13**, 41–42 (1929).

[2] E. C. Titchmarsh, *J. Lond. Math. Soc.*, **2**, 148–150 (1927).

[3] S. Bernstein, *C. R.*, **158**, 1661–1663 (1914).

[4] A. Zygmund, *J. Lond. Math. Soc.*, **3**, 194–196 (1928).

[5] The Fourier series analogue of this theorem is due to M. Riesz. See M. Riesz and E. Hilb, *Encykl. math. Wissensch.*, **2** (3), p. 1211; M. Riesz, *Math. Z.*, **27**, 218–244 (1927), especially p. 231; and G. H. Hardy and J. E. Littlewood, *J. Lond. Math. Soc.*, **3**, 250–253 (1928).

[6] The Fourier series analogue of this theorem is due to G. H. Hardy and J. E. Littlewood, *Math. Annalen*, **97**, 159–209 (1927). See also their paper quoted in the preceding footnote.

[7] E. Hille and J. D. Tamarkin, *Amer. Math. Monthly*, **36**, 255–264 (1929).

THE ARCHITECTURE OF LIVING CELLS—RECENT ADVANCES IN METHODS OF BIOLOGICAL RESEARCH—OPTICAL SECTIONING WITH THE ULTRA-VIOLET MICROSCOPE

BY FRANCIS F. LUCAS

BELL TELEPHONE LABORATORIES, INC., NEW YORK CITY

Read before the Academy April 28, 1930

In papers[1,2,3] presented elsewhere during the past few years, the development and application of the ultra-violet microscope to the science of metallography has been described.

Metallography, at first thought, appears wholly unrelated to histology or other branches of biology but the two branches of science do have many points in common. Both deal in the last analysis with the structure of matter and, in each, the microscope is an indispensable tool. Improvements in microscopic vision which enlarge the world of vision in one branch of science inevitably have a reflection in the other.

It is not the purpose in this paper to enter into a discussion of structures of living cells as revealed by the ultra-violet microscope. More particularly, the object is to present a tool for biological research; a tool which enables us to photograph the structure at different planes or levels within a single cell or group of cells; one which enables us to see the living cell with a degree of precision and clarity not heretofore possible by any other known means and with a potential resolving ability at least twice that of the best apochromatic system using visible light.

We are now able to take a series of photographs at slightly different focal planes or levels through a single cell or group of cells. These photographs result in a consecutive record of the internal architecture of the cell or cells. This accomplishment we have called "optical sectioning."

Nearly thirty years ago,[4] Dr. A. Köhler of the Zeiss Scientific Staff developed the ultra-violet microscope for use in the field of biology. It has a potential resolving ability twice that of the best optical system using visible light. The equipment was made available to scientists by the firm of Carl Zeiss and some sporadic work was done by different investigators but no outstanding accomplishment has appeared to its credit. Difficulty in manipulation and inability to focus the equipment seem to have reacted against the widespread use of the ultra-violet microscope.

Nearly ten years ago, following the development of high power metallography and the utilization of the full potential resolving ability of the best apochromatic systems using visible light, Bell Telephone Laboratories arranged with the Zeiss Scientific Staff to design an ultra-violet microscope suitable for metallography and for transparent specimens.

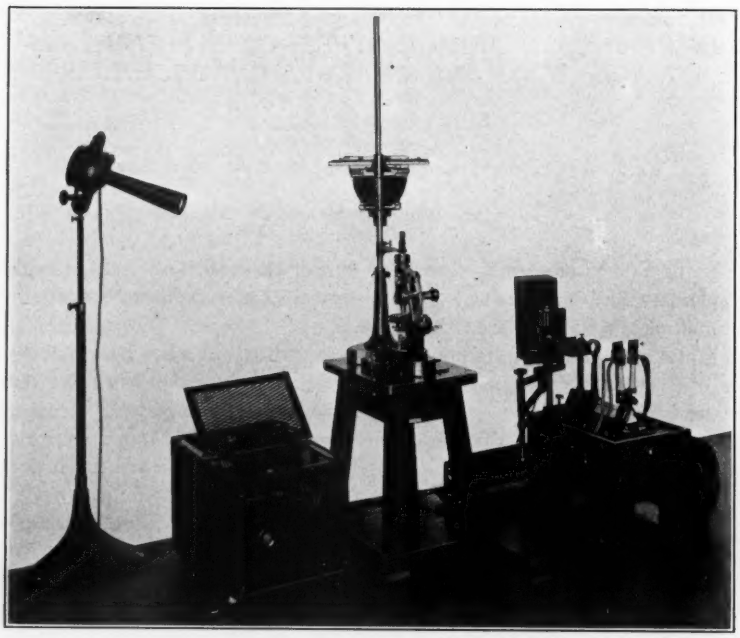


FIGURE 1

ULTRA-VIOLET MICROSCOPE ARRANGED FOR OPTICAL SECTIONING OF LIVING CELLS

The combined sound and fume hood is shown removed from the spark generating apparatus and the searcher eyepiece is in position above the ocular of the microscope.

The ultra-violet microscope is a monochromatic system corrected for wave length—275$\mu\mu$. The optical parts are of quartz. Monochromatic light is necessary due to lack of suitable optical materials with which to correct the objectives for chromatic differences. The source of illumination is a spark gap and a quartz slit. Cadmium electrodes are used. The potential at the electrodes is 10,000 volts secured from a primary source of 220 volts at 60 cycles by means of a step-up transformer. The light passes through a collimator and two 60° prisms to emerge from the prism or field diaphragm in the form of a line spectrum. The desired line (λ 2748.67) enters a prism at the base of the microscope and is directed upward into the substage condenser.

Since we are dealing with invisible light, the optical image cannot be seen in the ordinary way. It must either be photographed or visualized on some sort of a fluorescent screen. The fluorescent screen takes the form of an artificial eye, which is a uranium glass wedge incorporated in a small optical system and mounted interchangeably with a camera just above the ocular. The surface of the wedge on which the image is to be received has two cross lines. These cross lines are brought into focus by a small magnifier. Then the fluorescing image of the object is focused on the plane of the cross lines by the adjustments of the microscope in the usual way. When the image is once in focus in the artificial eye or searcher eyepiece as it is called, the calculations of the optical system are such that it will be in focus in a plane 30 centimeters above. The camera is then substituted for the searcher eyepiece and a photograph taken.

After working with the equipment as a metallurgical microscope, it soon became evident that a more exact method of focusing must be devised than the "searcher eyepiece" method of Köhler. In most cases it was impossible, except by pure chance, to secure an exact focus. The solution of this problem was discussed in a lecture before the American

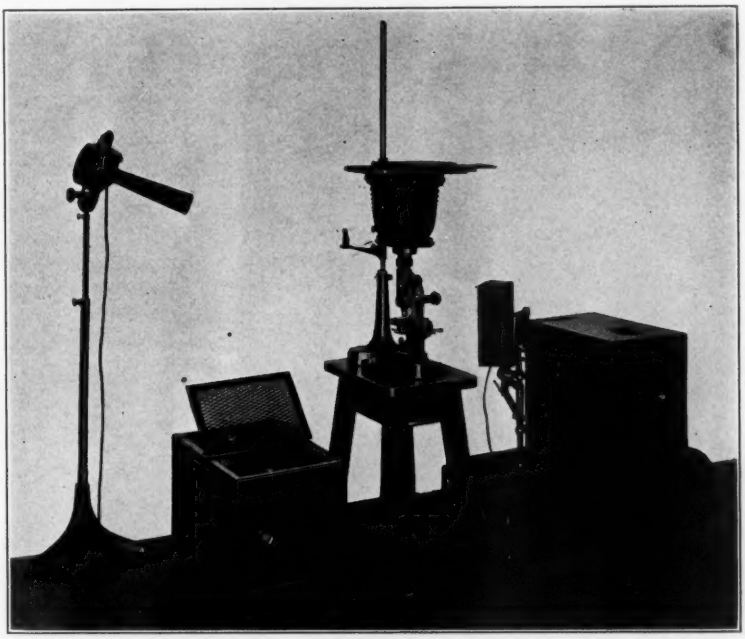


FIGURE 2

ULTRA-VIOLET MICROSCOPE ARRANGED TO PHOTOGRAPH OPTICAL SECTIONS

The sound and fume hood is in place. The light emerges through the open window shown. The window is fitted with glass and when swung closed excludes the ultra-violet light. In this way, the specimen can be protected from the ultra-violet light except during intervals of exposure. The fumes generated by the spark are carried away by an exhaust fan. The metal exhaust outlet, cushioned to prevent vibration, is shown at the extreme right of the spark generating apparatus. The camera is swung around to replace the searcher eyepiece.

Society of Mechanical Engineers and appears[3] also in published form to which those interested are referred. The details are briefly recounted here.

As a metallurgical microscope, it was found that the focus must be exact to at least one-quarter micron. The graduations on the drum of the slow motion adjustment of the microscope are by increments of one

or two microns, depending on the construction of the particular slow motion mechanism used on the microscope. It was necessary to increase the sensitivity of the fine adjustment. The peripheral motion of the drum was amplified by a pointer fitted to the thumb-screw of the fine adjust-

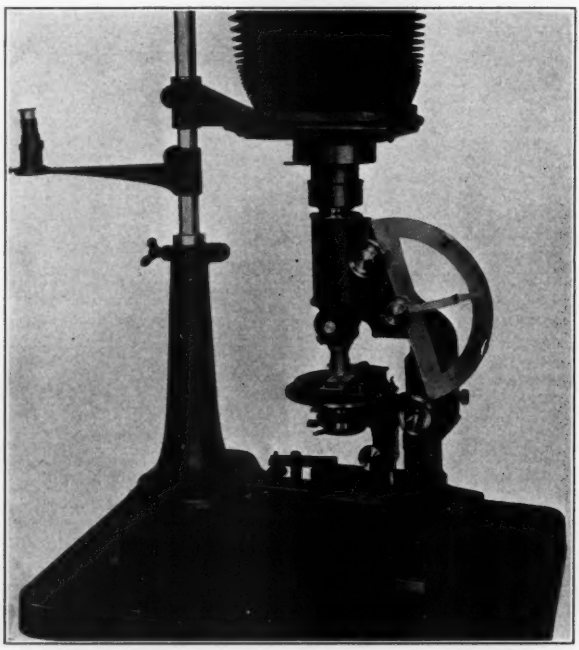
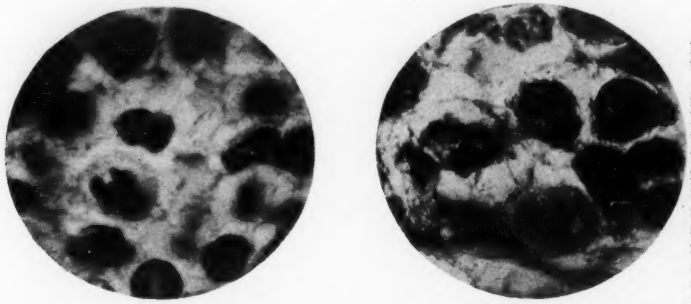


FIGURE 3

ULTRA-VIOLET MICROSCOPE

Showing the microscope fitted with the graduated half-circle, graduated slider and the aluminum pointer attached to the slow motion thumb-screw. In this particular assembly, one degree on the protractor corresponds to a change in focal planes of one-quarter micron. A sensitivity of one-half degree in adjustment may easily be attained. By means of a different slow motion mechanism, a spacing in focal planes of about one-sixteenth micron may be secured. A spacing of one-quarter micron seems adequate for most work.

ment. The pointer functioned with a graduated half-circle attached to the microscope stand. An ordinary draftsman's protractor was used for the graduated half-circle. The graduated rim of the protractor was fitted with a slider which also was graduated, each division equaling four degrees.



A B

FIGURE 6

A fixed and embedded specimen, carcinoma of the appendix, photographed by visual light methods and by the ultra-violet microscope.

The ultra-violet microscope has an enormous increase in resolving ability over the visual light system and much higher magnifications may be employed. Figure 6-A was taken with the best apochromatic system (NA 1.40) using visual light. The magnification is 1000×, reduced $^1/_2$ in reproduction. On this particular specimen, an actual loss in definition occurred at 1500×. Figure 6-B shows an unstained section from the same paraffin block photographed at 1500× with the ultra-violet microscope, reduced $^1/_2$ in reproduction. The sections were identical except one was stained and mounted on glass for the visual light microscope. The other was unstained and mounted on quartz for the ultra-violet microscope. Figure 6-B is one of a group of optical sections.

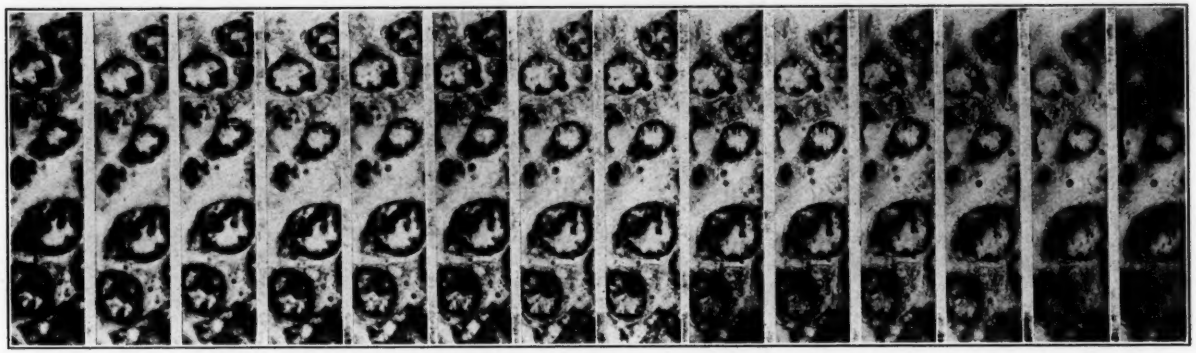


FIGURE 7

Optical sections of a fixed but unstained specimen. The specimen is a mouse tumor photographed on planes spaced one-quarter micron apart. The original magnification is 1800×, reduced $^1/_2$ in reproduction.

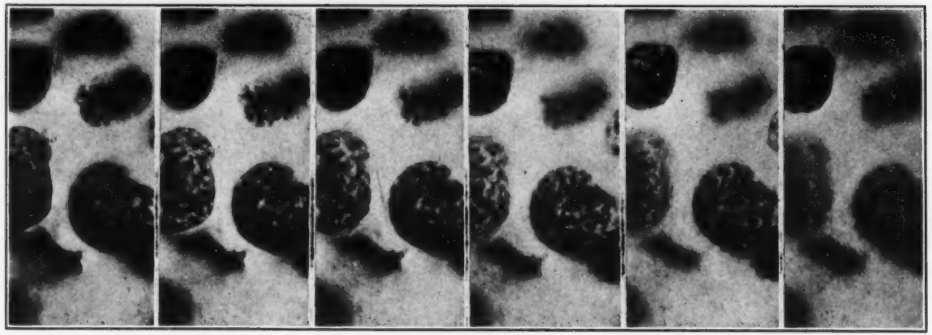


FIGURE 8

Optical sections of living mouse tumor cells. Neither fixative nor stain was used. The focal planes are spaced one-quarter micron apart. The magnification is 1800×, reduced $^1/_2$ in reproduction.

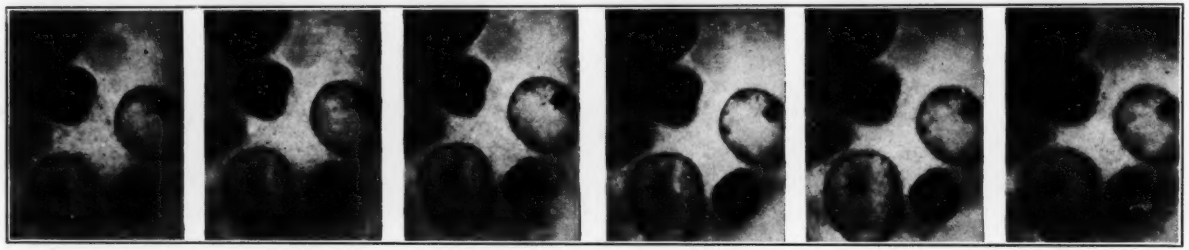


FIGURE 9

Six optical sections of mouse tumor cells. The focal plane spacing is one-quarter micron and the magnification is 1800 × reduced, $^1/_3$ in reproduction. Locke's solution was used as a mounting medium. No stain was used. Observe the progressive changes in the very fine structural details.

The method of operation is as follows: The image in the searcher eyepiece is brought into approximate focus by the usual methods. The slow motion is then turned so that the instrument passes downward through the focus and then it is very gradually brought back. In this way all lost motion in the slow adjustment is compensated for by gravity. A very nice mechanical balance must be obtained in the moving parts, otherwise, the method will fail. Any "creep" or change in focus, even though very slight, will upset the whole method. The personal element perhaps is still most important because the method relies on the keenness of vision of the operator to bring the image in the searcher eyepiece just

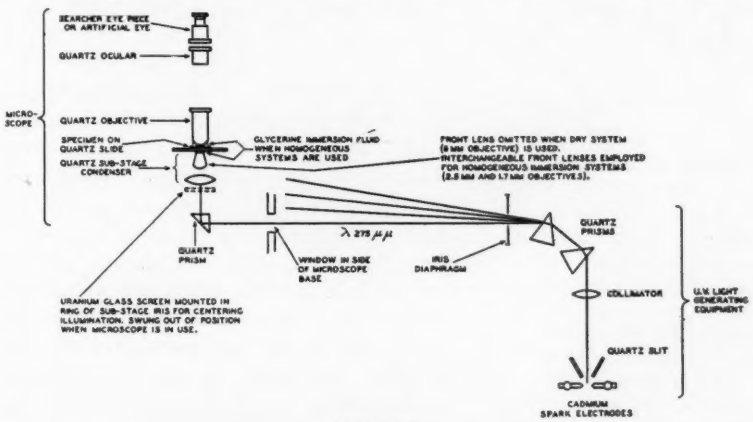

SEARCHER EYE PIECE OR ARTIFICIAL EYE
QUARTZ OCULAR
MICRO-SCOPE
QUARTZ OBJECTIVE
SPECIMEN ON QUARTZ SLIDE
QUARTZ SUB-STAGE CONDENSER
GLYCERINE IMMERSION FLUID WHEN HOMOGENEOUS SYSTEMS ARE USED
FRONT LENS OMITTED WHEN DRY SYSTEM (8 MM OBJECTIVE) IS USED. INTERCHANGEABLE FRONT LENSES EMPLOYED FOR HOMOGENEOUS IMMERSION SYSTEMS (2.5 MM AND 1.7 MM OBJECTIVES).
λ 275 μμ
QUARTZ PRISMS
QUARTZ PRISM
WINDOW IN SIDE OF MICROSCOPE BASE
IRIS DIAPHRAGM
URANIUM GLASS SCREEN MOUNTED IN RING OF SUB-STAGE IRIS FOR CENTERING ILLUMINATION. SWUNG OUT OF POSITION WHEN MICROSCOPE IS IN USE.
COLLIMATOR
U.V. LIGHT GENERATING EQUIPMENT
QUARTZ SLIT
CADMIUM SPARK ELECTRODES

FIGURE 4

Schematic diagram of optical systems of the ultra-violet microscope as arranged for transparent specimens.

to the stage of coming into focus. The pointer is then set near the middle of the scale and the slider moved so that the first graduation coincides with the pointer. A photograph is taken with the apparatus in this position. The pointer is then lowered one degree which corresponds to a change of one-quarter micron in elevation of focus, the mechanical arrangements of the parts being arranged accordingly and a second photograph taken and so on. Four photographs taken in this way of a metallurgical specimen are sufficient to provide one photograph in exact focus.

Any optical system which will yield a very sharply defined focal plane to an accuracy of about one-quarter micron should prove to be of great value. For example, it appeared reasonable to suppose that we might substitute a transparent preparation for the opaque metallurgical specimen, and, because of the transparency of the specimen, we should be

able to photograph details of structure on different planes. Obviously if the focus is confined to a focal plane of inappreciable depth, detail above or below the exact focal plane should not interfere with the image.

From the solution of one problem dealing with the focus of a metallurgical microscope has evolved a method of optically sectioning living cells.

Dr. E. E. Free has written of this development by likening a single cell to a house and he has imagined what he would see if the house were divided by equally spaced cutting planes. I shall adopt Dr. Free's analogy to make clear what actually does happen when living cells are so sectioned.

FIGURE 5

Optical sectioning with the ultra-violet microscope

By means of this new development a transparent specimen such as a group of cells may be sectioned optically, the focal planes may be spaced as closely as $^1/_{16}$ micron—generally a spacing of $^1/_4$ micron suffices. Detail above or below the focal plane does not interfere. Successive photographs taken on planes A, B, C, etc., give a progressive record of the structure.

Magnifications as high as 5000 diameters result in crisp brilliant images with a degree of resolution (ability to reveal fine detail) surpassing by far that achieved with any other known optical system.

First we may take a photograph on a plane cutting through the attic. We do not see the roof above or the floors below. We see only the details on the exact focal plane selected. Perhaps there will be trunks, bags, boxes and things of another day placed in the attic for storage. All of these things on the focal plane are clearly defined. Perhaps our focal plane cuts midway through a trunk; we do not see the top or bottom of the trunk but we see the things which are stored away within the trunk and which are intersected by the cutting focal plane. In one compartment of the trunk, our focal plane has come on some woolens. We see the texture of the cloth and observe that it shows wear and then we come on the handiwork of the moth larvæ. In another compartment are some queer-looking sections of straw and cloth and wire and we decide that it must be a hat on which the sun has risen and no doubt set for the last time. And so we could investigate the contents of all the trunks, bags and boxes without opening them, disturbing their contents or doing the slightest damage.

The next optical section intersects the sleeping and bathrooms of the second floor. At a glance, we note the number and arrangement of the rooms. We recognize the furniture and its disposition about the room. One room we conclude is a nursery because children are at play with things which we recognize as toys and the play and movement goes along entirely oblivious of the scrutiny from above. In another room, our optical section coincides with the top of a dressing table. We see all of

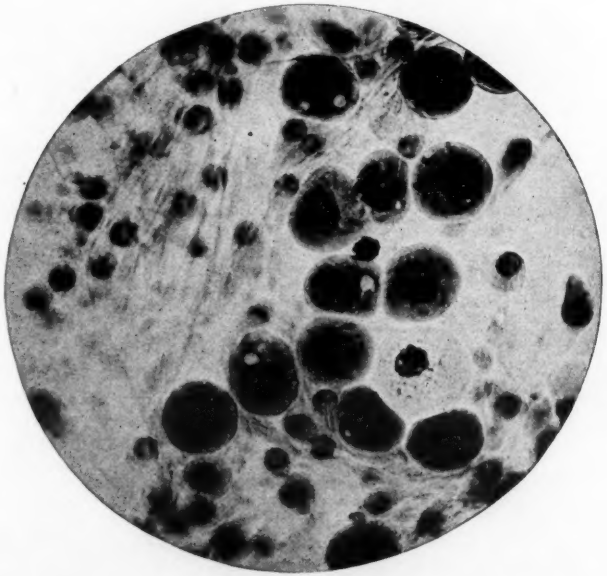


FIGURE 10

Living sperm cells of the grasshopper, photographed with the ultra-violet microscope. One of a series of optical sections. Ringer's Solution was used as a mounting medium. No stain was used. The magnification is 500×, reduced 1/3 in reproduction.

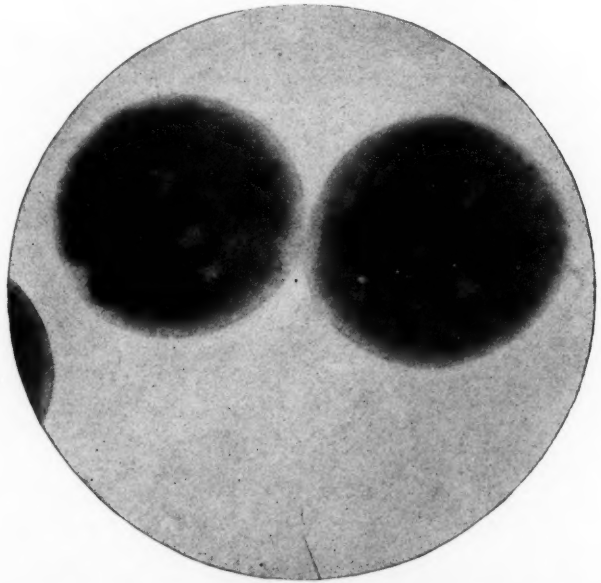


FIGURE 11

Individual cells of the same specimen as Figure 10. Magnification is 1800×, reduced 1/3 in reproduction.

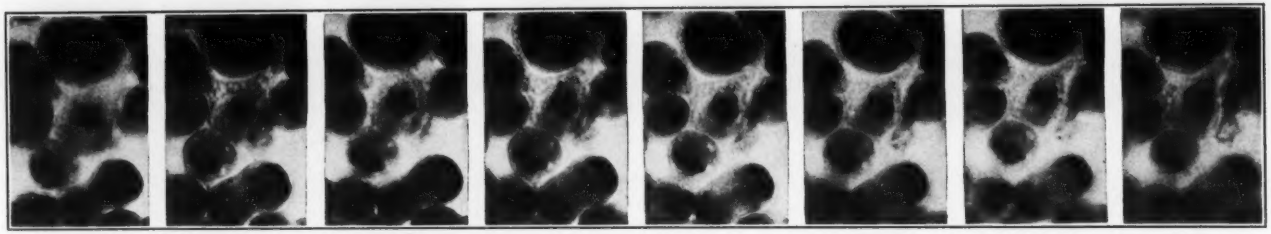


FIGURE 12-A

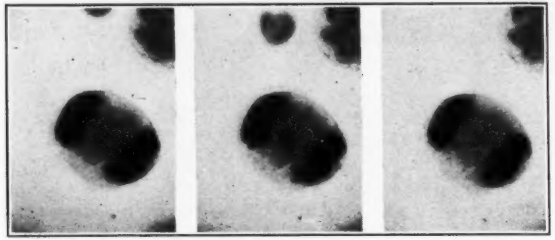


FIGURE 12-B

Figures 12-A and 12-B. Living sperm cells of the grasshopper. The focal planes are spaced one-quarter micron. The magnification is 1800×. Figure 12-A has been reduced $^2/_3$ in reproduction, 12-B reduced $^1/_2$. Note absorption of light by the chromosomes and absence of spindle fibers. The preparation was mounted in Ringer's Solution. No stain was used.

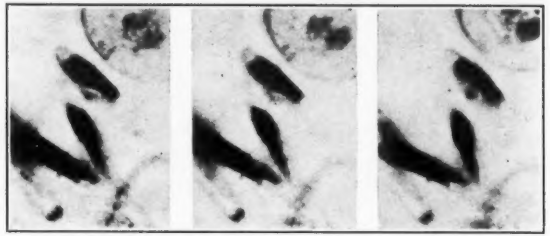


FIGURE 13

Living unstained spermatozoa of the mouse. The optical sections are spaced one-quarter micron. The magnification is 3600×, reduced $^1/_2$ in reproduction.

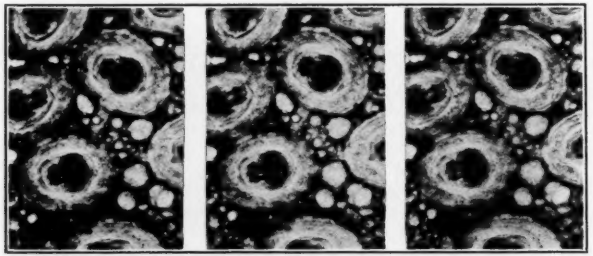


FIGURE 14-A

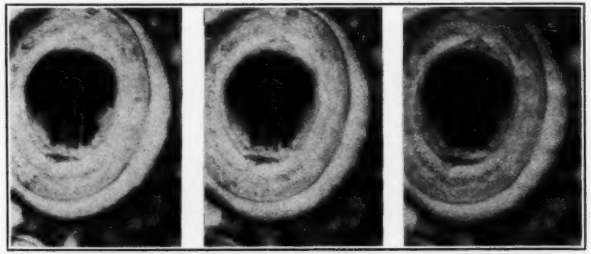


FIGURE 14-B

Embryos within the segments of a small tapeworm found within the body cavity of a mouse. Identification by Dr. Stark.
Figure 14-A shows three optical sections at 500× and Figure 14-B three optical sections at 1200×, both reduced 1/3 in reproduction. The focal plane spacing is one-quarter micron. The preparation was mounted in Locke's Solution and was unstained.

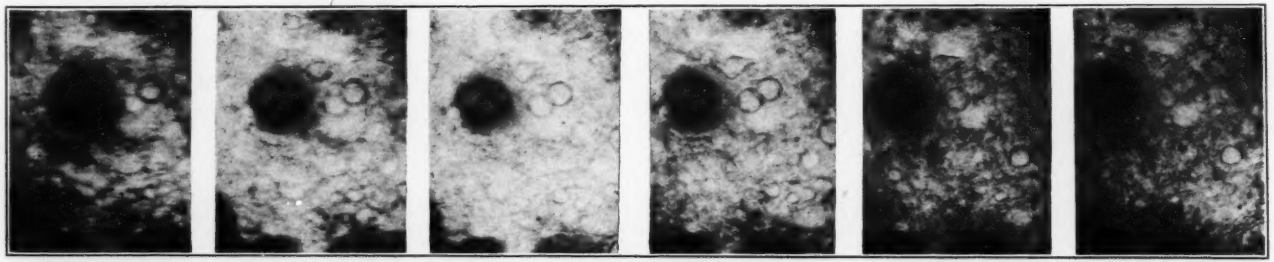


FIGURE 15-A

Golgi cells of the cerebellum of the mouse. Magnification is 1800×, reduced 1/3 in reproduction. Cells mounted in Ringer's Solution with faint neutral red stain. Optical spacing one-quarter micron.

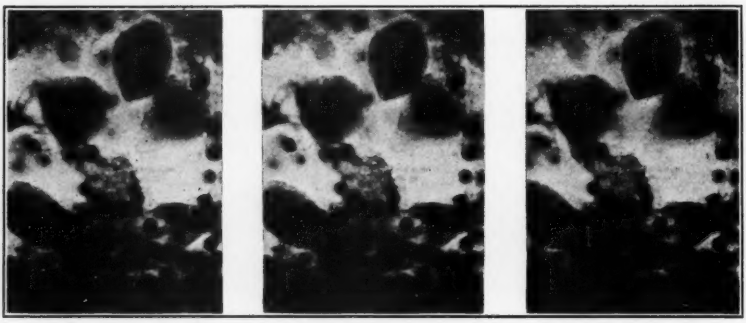


FIGURE 15-B

Purkinje cells of the cerebellum of the guinea pig. Magnification is 500×, reduced 1/4 in reproduction. Cells mounted in Ringer's Solution with faint neutral red stain. Optical spacing one-quarter micron.

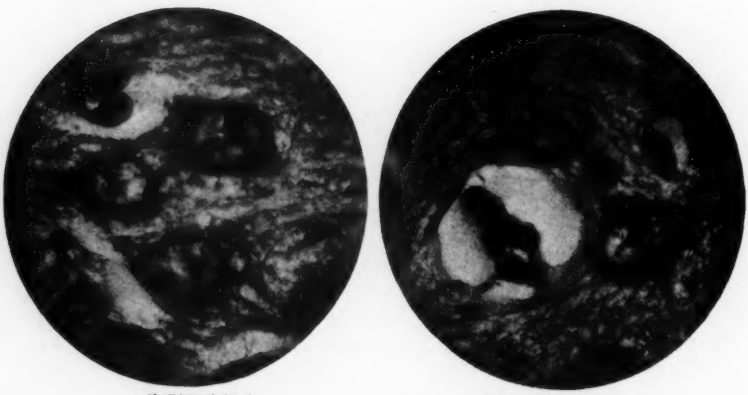


FIGURE 16-A

FIGURE 16-B

Cerebellum of the cat. The specimen was fixed but not stained. Illustrating the application of the ultra-violet microscope to the study of fixed but unstained brain cell structures. Original magnification is 1800×, reduced 1/2 in reproduction.

the toilet articles and they have apparently been put down in haste as things are in disorder. In a jewel case cut by the sectioning plane, are many pieces of jewelry which we pause to admire. A large diamond ring which has been itself sectioned strikes our fancy because we look within and see a black carbon spot which we surmise the owner does not know is present. We discover something wrong with the contour of the stone. It is altogether too broad for its depth and we call it a "fish eye" and pass on.

And so we may go from room to room and from floor to floor observing the life and habits of the occupants. If we take photographs as we go, they may be laid out in order from top to bottom and a working model of the house constructed.

Fortunately, our private lives are not open to such scrutiny as the hypothetical example just cited. Though it does describe on a large scale what we are now able to do with many living things. A transparent specimen such as a single living cell or a group of living cells may be sectioned optically. The focal planes may be spaced as closely as $^1/_4$ micron. Detail above or below the focal plane does not interfere. Photographs taken on successive planes give a progressive record of the structure within the cell.

Dr. Köhler pointed out many years ago that one advantage of the ultra-violet microscope lay in the fact that organic specimens are differentiated in structure by virtue of the selective absorption which they manifest toward ultra-violet light. When using visual light microscopes, the need for staining biological specimens is too well known to need comment here. Unstained specimens respond under the ultra-violet microscope much as though they were stained.

It is generally recognized that the structure of organic material is apt to be profoundly altered by the treatment in preparing it for microscopic examination. The trend in cytological research appears to be toward the study of living material thus avoiding artifacts induced by fixation, staining and mounting. In the final analysis, biologists are interested in the structure, functions and behavior of the living undisturbed cell.

It may be argued that any cell removed from its normal living habitat and placed in some artificial medium is no longer a normal cell though it may be a living cell. In some cases, well justified doubts may be entertained as to whether the cell is actually alive. By what criteria shall we decide matters in this borderland? Shall we answer by opinion or careful laboratory observations or by both? Ultra-violet microscopy of living cells is in its infancy. We have observed by experimental methods on fresh material and aged material that changes do occur. In some cases the material undergoes quite rapid changes. In other cases, a surprisingly long interval may elapse before any change or degeneration can be de-

tected. Cells of the nervous system are known to be most delicate of all and we have found them more difficult to work with than other types of cells. With respect to cells of the nervous system, our results have not been fully studied, nevertheless the photographs perhaps will be found of interest since they are believed to represent at least a close approximation to normal.

If we are to employ the ultra-violet microscope in such researches, there are two reservations which must be borne in mind. The ultra-violet light itself must not have apparent injurious effects on the organisms and the organisms must not completely absorb the ultra-violet light. Things which completely absorb the ultra-violet light cannot be studied by these methods.

Ultra-violet light of the intensity and wave length used in conjunction with the ultra-violet microscope appears to have little if any harmful effect so far as some types of living cells are concerned. It is true that some single cell organisms are destroyed almost instantly; others of another species in the same mixed culture appear mildly excited and others immune. From some preliminary observations, it appears that living cell cultures may be exposed under the ultra-violet microscope for as long as forty-five minutes and, when returned to the incubator, they appear to grow and to show no ill effects from the exposure.

One wonders what effects wave-lengths other than $275\mu\mu$ would have on certain standard cell cultures. Will monochromatic ultra-violet light of one wave-length cause a living cell to disintegrate immediately and similar light of another wave-length have little or no effect? If so, then a very powerful means of dealing with cells will be at hand.

To illustrate the application of the methods used, I have chosen a variety of cell structures. The tissues were taken from the animals and transferred to physiological salt solutions. Cell smears were quickly made on quartz slides; a drop of the salt solution added and the preparation covered with a quartz cover slip. The cover slip was sealed with vaseline to prevent evaporation from the preparation. The salt solutions were chosen with particular regard to their suitability for the purpose, the object being to maintain isotonic conditions with reference to the cell. In some cases we have used the body fluids of the cells themselves instead of the artificial salt solutions.

It will be observed that the results with the ultra-violet microscope often show marked departures from corresponding structures photographed with visual light systems. This is particularly noticeable not alone because of the wealth of detail revealed by ultra-violet methods but also because of the apparent lack of some details. An example of this is found in some mitotic figures where the spindle fibers appear wanting or only faintly suggested by ultra-violet methods.

In the preparation of the material and the interpretation of the results, I have had the coöperation of Dr. Mary B. Stark of the Flower Hospital Staff and the assistance of Miss A. K. Marshall of the Bell Telephone Laboratories.

The photographs which accompany this paper are sufficiently described in the titles and need not be specifically dealt with here.

[1] Lucas, "An Introduction to Ultra-Violet Metallography," *Trans. Am. Inst. Mining and Metallurg. Eng.*, February, 1926.

[2] Lucas, "A Resumé of the Development and Application of High Power Metallography and the Ultra-Violet Microscope," **1**, *Proc. Int. Cong. Test. Materials*, Amsterdam, Holland, 1927.

[3] Lucas, "Photomicrography and Its Application to Mechanical Engineering," *Mech. Eng.*, **50**, pp. 205–212, March, 1928.

[4] Köhler, "Microphotographic Examinations with Ultra-Violet Light," *Zeit. Wissensch. Mikroskopie und für Mik. Tech.*, **21**, 1904, pp. 129–165 and 273–304.

A FUNDAMENTAL THEOREM ON ONE-PARAMETER CONTINUOUS GROUPS OF PROJECTIVE FUNCTIONAL TRANSFORMATIONS[1]

BY L. S. KENNISON

DEPARTMENT OF MATHEMATICS, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated August 11, 1930

Let L^x, L^x_s, L^x_1, L^1_s be real bounded integrable functions of the real variables x and s as indicated ($a \leq x, s \leq b$) and let us denote Riemann integration on (a, b) by the repetition of a superscript and subscript in the same term unless one of them is enclosed in a parenthesis.

The regular infinitesimal projective transformation in function space

$$\varphi^x = \bar{\varphi}^x + \delta t[L^x\bar{\varphi}^x + L^x_s\bar{\varphi}^s + L^x_1 - \bar{\varphi}^x L^1_s\bar{\varphi}^s] \tag{1}$$

will generate by continuous application a family of projective functional transformations[2]

$$\bar{\varphi}^x(t) = \frac{K^x(t)\varphi^x + K^x_s(t)\varphi^s + K^x_1(t)}{K^1_s(t)\varphi^s + K^1_1(t)} \tag{2}$$

where $\bar{\varphi}^x(t)$ satisfies the integro-differential system

$$\left.\begin{aligned} \frac{\partial\bar{\varphi}^x(t)}{\partial t} &= L^x\bar{\varphi}^x(t) + L^x_s\bar{\varphi}^s(t) + L^x_1 - \bar{\varphi}^x(t)L^1_s\bar{\varphi}^s(t) \\ \bar{\varphi}^x(0) &= \varphi^x \end{aligned}\right\} \tag{3}$$

Dines[3] has shown that in order for $\bar{\varphi}^x(t)$ in (2) to satisfy (3) the following equations hold

$$\left.\begin{aligned} 'K_s^x &= L^x K_s^x + L_s^x K^{(s)} + L_u^x K_s^u + L_1^x K_s^1 \\ 'K_1^x &= L^x K_1^x + L_u^x K_1^u + L_1^x K_1^1 \\ 'K_s^1 &= L_s^1 K^{(s)} + L_u^1 K_s^u \\ 'K_1^1 &= L_u^1 K_1^u \\ 'K^x &= L^x K^x \end{aligned}\right\} \tag{4}$$

$$K^x(0) = K_1^1(0) = 1, \qquad K_s^x(0) = K_1^x(0) = K_s^1(0) = 0, \tag{5}$$

where the primes indicate differentiation with respect to t.

If (α, β) represent any set of indices which the K's may have, assume a solution of (4) and (5) as follows

$$K_\beta^\alpha(t) = \sum_{m=0}^{\infty} \frac{t^m}{m!} K_\beta^\alpha [m] \tag{6}$$

Substituting in (4) and equating coefficients of powers of t, we obtain the recurrence formulas determining the $K_\beta^\alpha[m]$ successively

$$\begin{aligned} K_s^x[m+1] &= L^x K_s^x[m] + L_s^x K^{(s)}[m] + L_u^x K_s^u[m] + L_1^x K_s^1[m] \\ K_1^x[m+1] &= L^x K_1^x[m] + L_u^x K_1^u[m] + L_1^x K_1^1[m] \\ K_s^1[m+1] &= L_s^1 K^{(s)}[m] + L_u^1 K_s^u[m] \\ K_1^1[m+1] &= L_u^1 K_1^u[m] \\ K^x[m+1] &= L^x K^x[m] \end{aligned} \tag{7}$$

and from (5) we have

$$K^x[0] = K_1^1[0] = 1, \quad K_s^x[0] = K_1^x[0] = K_s^1[0] = 0. \tag{8}$$

A dominating series is readily found so that the formal solution (6) is an actual solution of (4) and (5). Hence the transformations (2) generated by (1) from a one-parameter analytic family.

The following recursion formulas will be needed

$$\begin{aligned} K_s^x[p] &= K^x[m]\, K_s^x[p-m] + K_s^x[m] K^{(s)}[p-m] + K_u^x[m] \\ &\quad K_s^u[p-m] + K_1^x[m]\, K_s^1[p-m] \\ K_1^x[p] &= K^x[m]\, K_1^x[p-m] + K_u^x[m]\, K_1^u[p-m] + K_1^x[m] K_1^1[p-m] \\ K_s^1[p] &= K_s^1[m]\, K^{(s)}[p-m] + K_u^1[m] K_s^u[p-m] + K_1^1[m]\, K_s^1[p-m] \\ K_1^1[p] &= K_u^1[m]\, K_1^u[p-m] + K_1^1[m]\, K_1^1[p-m] \\ K^x[p] &= K^x[m]\, K^x[p-m] \end{aligned} \tag{9}$$

which we shall abbreviate

$$K_\beta^\alpha[p] = K_\gamma^\alpha[m]\, K_\beta^\gamma[p-m] \tag{10}$$

For $(p, m) = (0, 0), (1, 0), (1, 1)$ these are easily verified. The general case is proved by induction on p.

It is easily verified that the kernels of the product of two transformations of type (2) generated by the same infinitesimal transformation (1), with parameters t_1 and t_2 is given by

$$P^{\alpha}_{\beta}(t_1,t_2) = K^{\alpha}_{\gamma}(t_1)K^{\gamma}_{\beta}(t_2) \qquad (11)$$

where the Greek indices take on exactly the same values as in (10). This is the crux of the proof of the theorem below.

THEOREM. *The one-parameter family of projective functional transformations* (2) *generated by a regular infinitesimal projective functional transformation* (1) *is a one-parameter continuous group.*

[1] A general theory of linear functional equations on a composite range with application to projective functional transformations including a fuller account of the work of this note is to be published elsewhere. These developments are embodied in a California Institute thesis. I am indebted to Prof. A. D. Michal for suggesting these topics and for invaluable suggestions and criticisms.

[2] L. L. Dines, *Trans. Am. Math. Soc.*, **20**, 45 (1919), has given in different notation the inversion and group properties for transformation of type (2) and has shown the existence of the one-parameter family satisfying (4) and (5). G. Kowalewski, *Wien. Ber.*, **120**, 1435, has given the name "regular infinitesimal projective functional transformation" to (1).

[3] Loc. cit., p. 59; see also, I. A. Barnett, *Bull. Am. Math. Soc.*, **36**, 273 (1930)

A SPECIAL TYPE OF UPPER SEMI-CONTINUOUS COLLECTION[1]

BY HARRY MERRILL GEHMAN

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF BUFFALO

Communicated July 26, 1930

1. *Introduction.*—The object of this paper is to show, in the final section, an application of the special type of upper semi-continuous collection of continua[2] which is discussed in § 3. Before doing so, we shall prove certain theorems concerning upper semi-continuous collections in general.

2. *G-Maps on a Cactoid.*—R. L. Moore has shown that an upper semi-continuous collection of continua which fills up a sphere is topologically equivalent to a cactoid.[3] Since a plane is topologically equivalent to a sphere minus a point, this theorem can be extended to the case where the collection fills up a plane, in which case the collection is topologically equivalent to a cactoid minus a non-cut point.

If then G is an upper semi-continuous collection of continua which fills up a sphere (or plane) S, a given correspondence T between the elements of G and the points of a cactoid (or cactoid minus a non-cut point) Σ, affects a kind of "mapping" of the points of S upon the points of Σ. To define this "mapping" more precisely:

Let F be any subset of S, and let G_F be the collection of elements of G

which have points in common with F. If Φ is the set of points of Σ, which correspond to the elements of G_F under the correspondence T, then we shall call the set Φ the *G-map of F on* Σ. It should be noted that the G-map of a set F depends upon the correspondence T as well as upon the collection G.

If the set F has a point in common with an element X of the collection G, then it is obvious that the sets F and $F + X$ will have on Σ the same G-map Φ. In general, we have:

THEOREM 1. *If H_F is the set of all points belonging to the elements of G_F, and if K is any subset of H_F, then the sets F and $F + K$ will have on Σ the same G-map.*

THEOREM 2. *The G-map of a continuum is a continuum in Σ.*

For if F is a continuum, then H_F is a continuum,[4] and hence the G-map of F on Σ is a continuum in Σ—that is, is connected and closed in Σ.

THEOREM 3. *The G-map of a continuous curve is a continuous curve in Σ.*

This theorem follows at once from Theorem 2 and the properties of continuous curves and of upper semi-continuous collections.

As before, let H_F denote the set of all points belonging to the elements of G_F, where G_F is the collection of elements of G which have points in common with F. Let H_A denote the set of all points of H_F which lie in elements of G_F consisting of a single point, and let $H_B = H_F - H_A$. By definition, H_A is a subset of F.

THEOREM 4. *If the G-map on Σ of a continuum F is a continuous curve, then F is connected im kleinen at every point of H_A which is not a limit point of H_B.*

Suppose on the contrary that a continuum F is not connected im kleinen at some point P of H_A which is not a limit point of H_B, and that Φ, the G-map of F on Σ, is a continuous curve. Then within some neighborhood R of P that contains no points of H_B, we have the state of affairs described in § 3 of R. L. Moore's "Report on Continuous Curves."[5] Since there is a continuous (1–1) correspondence between the points of F in the neighborhood R, and the points of some subset of Φ, it follows that Φ fails to be connected im kleinen at some point, which is contrary to hypothesis.

The following example shows the necessity of assuming that the point P is not a limit point of the set H_B.

Example.—In a Euclidean 3-space, let S be the sphere $x^2 + y^2 + z^2 = 1$. Let the elements of G be the circles in which the planes parallel to the XY-plane intersect the sphere. The collection G is topologically equivalent to the straight line from (0, 0, 1) to (0, 0, −1). Since every subcontinuum of a straight line is a point or a straight line, it follows that the G-map of *every* subcontinuum of S is a continuous curve. It should be noted, however, that for any subcontinuum F of S, the only points of S which could possibly be points of H_A are the points (0, 0, 1) and (0, 0, −1),

and if F contains one of these points, then that point is a limit point of H_B.

3. *A Special Type of G-Map.*—Let L be a bounded continuum lying on a sphere S. The collection of continua G_L consisting of L and of the points of S-L is an upper semi-continuous collection and is topologically equivalent to the cactoid Σ consisting of a contracting sequence of spheres, each one of which is tangent to each one of the other spheres at a point Λ. Such a set has been called by R. L. Moore a simple aspiculate cactoid.[6]

In case the continuum L lies in a plane S, the surface Σ may be defined as the cactoid just described, with some point (different from Λ) omitted from one of the spheres, or it may be defined as a surface consisting of a plane Π and a contracting sequence of spheres, each one of which is tangent to the plane Π at the same point Λ of the plane. In the discussion of § 4, we shall consider Σ as the latter type of set.

THEOREM 5. *A necessary and sufficient condition that the G_L-map on Σ of a continuum F be a continuous curve, is that F be connected im kleinen at every point of S-L.*

The condition is necessary by Theorem 4, since $H_A = F(S-L)$ and $H_B = FL$ is closed. It is sufficient, because if Φ, the G_L-map of F, is not a continuous curve, it fails to be connected im kleinen at some point of $\Phi - \Lambda$, and then by an argument similar to that used in the proof of Theorem 4, it follows that F fails to be connected im kleinen at some point which is not a point of L.

THEOREM 6. *If L_1 and L_2 are bounded subcontinua of S and $L_1L_2 = 0$, a necessary and sufficient condition that a continuum F be a continuous curve is that the G_{L_1}-map of F on Σ_1 and the G_{L_2}-map of F on Σ_2 be continuous curves.*

The condition is necessary by Theorem 3. It is sufficient by Theorem 5, since under our hypotheses F is connected im kleinen at every point of $(S-L_1) + (S-L_2) = S$.

4. *An Application to Theorems on Accessibility.*—The above considerations can be used to show that the generalization of the concept of a *point's* being accessible by arcs (or by continua) to the concept of a *continuum's* being accessible by arcs (or by continua) which was recently announced by G. T. Whyburn[7] is a generalization which is more apparent than real.

As in § 3, let L be a bounded continuum lying in a plane S, and let G_L be the upper semi-continuous collection and Σ the surface described there. According to Whyburn's definition, L is said to be *accessible by arcs* from a point set D, provided that if A is any point of D, then G_L contains a simple continuous arc of elements from A to L, every point of which is contained in $D + L$. This is equivalent to saying that L is *accessible by arcs* from D, provided that if A is any point of Δ, the G_L-

map of D on Σ, then Σ contains a simple continuous arc from A to Λ, every point of which is contained in $\Delta + \Lambda$. That is, Whyburn's "generalized accessibility" in S is equivalent to "ordinary accessibility" in Σ.

Consideration of the G_L-map of D rather than D itself has the advantage that we replace an arc of elements, which when considered as a point set in S may not be an arc or even a continuous curve, by an ordinary arc of points in Σ. If then throughout the discussions in Whyburn's paper we replace each subset of S by its G_L-map on Σ, this replacing has the effect (1) of reducing his "generalization" to the ordinary notion of accessibility by arcs, and (2) of replacing the plane S by the more complicated space Σ. We shall next show that the difficulties involved in (2) are not sufficient to cause ordinary accessibility in Σ to be considered as a generalization of ordinary accessibility in S.

Methods of proof will not be materially changed if the space considered is Σ instead of S. So far as the topology of Σ is concerned, it is locally 2-dimensional, except in the neighborhood of the point Λ. Hence practically all topological theorems, particularly those concerning accessibility, hold true in this space, with appropriate modifications in some cases. If, for instance, L is a subset of a continuous curve F in S, the G_L-map of F on Σ is a continuous curve containing the point Λ, by Theorem 3. This continuous curve on Σ can be thought of as the sum of a sequence of continuous curves, one lying on the plane Π and one (which may degenerate into the point Λ only) on each of the spheres. Also if a set D is a subset of one component of S-L, then Δ the G_L-map of D on Σ lies on Π or on one of the spheres. Hence L is accessible from D if and only if the point Λ is accessible from a set Δ lying on a plane or a sphere. If D has points in common with more than one component of S-L, then L is accessible from D if and only if the point Λ is accessible from each one of a sequence of sets $\Delta_1, \Delta_2, \Delta_3, \ldots$, which lie respectively in a sequence of different spaces, one of which is a plane and the remainder are spheres.

As has just been shown, in considering questions of accessibility, it is in general no more difficult to prove theorems for the space Σ than for the space S. Hence it can be seen that the announced generalization of the idea of accessibility is in reality no generalization at all. Many of the results of Whyburn's paper[7] are merely restatements for the space Σ of theorems previously proved by Whyburn for the plane.[8]

While the above discussion has been restricted to the plane for the sake of definiteness, the same ideas can obviously be used in considering questions of accessibility in any Euclidean space of n dimensions.

[1] The major portion of this paper was presented to the American Mathematical Society at Bethlehem, Penna., December 27, 1929.

[2] For definitions, see: R. L. Moore, "Concerning Upper Semi-continuous Collections of Continua," *Trans. Amer. Math. Soc.*, **27**, 416–28 (1925).

[3] R. L. Moore, "Concerning Upper Semi-continuous Collections," *Monatshefte für Mathematik und Physik*, **36**, 81–88 (1929). See especially Theorem 2.

[4] R. L. Moore, *Transactions*, loc. cit., Theorem 1.

[5] *Bull. Amer. Math. Soc.*, **29**, 289–302 (1923).

[6] R. L. Moore, *Monatshefte*, loc. cit., Definitions, pp. 81–82 and Theorem 6. The set described above does not satisfy Moore's condition (*b*), p. 81—a condition which is not essential under any circumstances.

[7] G. T. Whyburn, "A Generalized Notion of Accessibility," *Fund. Math.*, **14**, 311–326 (1929).

[8] G. T. Whyburn, "On Certain Accessible Points of Plane Continua," *Monatshefte für Mathematik und Physik*, **35**, 289–304 (1928).

ON EXPANSIONS OF ARITHMETICAL FUNCTIONS

BY R. D. CARMICHAEL

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS

Communicated July 24, 1930

1. The remarkable expansions of arithmetical functions obtained by Ramanujan in his notable memoir (*Collected Papers*, No. 21) on certain trigonometrical sums and their applications are contained as special cases of much more general expansions which have also other special cases of particular interest. The purpose of this paper is to present these generalizations and to draw from the expansions some conclusions of importance obtained by means of a hitherto unnoticed fundamental property of the Ramanujan sums $c_q(n)$, namely, that expressed by the relations

$$\sum_{n=1}^{q\rho} c_q(n)c_\rho(n) = 0 \text{ if } \rho \neq q, \qquad \sum_{n=1}^{q} c_q^2(n) = q\varphi(q),$$

where $\varphi(q)$ denotes Euler's φ-function of q. This has led to the notion of orthogonal arithmetical functions analogous to the notion of orthogonal functions in analysis.

2. By $\chi(a)$, $\chi_1(a)$, $\chi_2(a)$, ..., we denote any characters modulis k, k_1, k_2, ..., respectively, and by $\chi_0(a)$, $\chi_{10}(a)$, $\chi_{20}(a)$, ... we denote the principal characters for these moduli. If a symbol for a character is written with an argument which is not an integer it may have for this argument any conveniently assigned value. The character which is equal to unity for all arguments will sometimes be replaced by 1. We use $\mu(a)$ to denote the Möbius function. By $\eta_q(n)$ we denote the sum of the n^{th} powers of the q^{th} roots of unity; then $\eta_q(n)$ is q or 0 according as q is or is not a factor of n.

We introduce the functions

$$c_q^{(\lambda)}(n, \chi, \chi_1, \chi_2) = \sum_{d/q} \mu\left(\frac{q}{d}\right) \chi\left(\frac{q}{d}\right) \chi_1(d)\chi_2\left(\frac{n}{d}\right) \eta_d^{\lambda}(n)$$

where the sum is taken for the divisors d of q and where λ is positive. This reduces to Ramanujan's function $c_q(n)$ when $\lambda = 1$ and $\chi(a) \equiv \chi_1(a) \equiv \chi_2(a) \equiv 1$.

Let $F(u, v)$ denote a function of u and v and write

$$D(n, \chi_1, \chi_2) = \sum_{d/n} \chi_1(d)\chi_2\left(\frac{n}{d}\right) F\left(d, \frac{n}{d}\right).$$

Then we have the following expansion:

$$D(n, \chi_1, \chi_2) = c_1^{(\lambda)}(n, \chi, \chi_1, \chi_2) \sum_{v=1}^{t} \frac{1}{v^{\lambda}} \chi(v) F\left(v, \frac{n}{v}\right)$$
$$+ c_2^{(\lambda)}(n, \chi, \chi_1, \chi_2) \sum_{v=1}^{1/2\,t} \frac{1}{(2v)^{\lambda}} \chi(v) F\left(2v, \frac{n}{2v}\right)$$
$$+ c_3^{(\lambda)}(n, \chi, \chi_1, \chi_2) \sum_{v=1}^{1/3\,t} \frac{1}{(3v)^{\lambda}} \chi(v) F\left(3v, \frac{n}{3v}\right) + \ldots, t \geqq n.$$

3. If we take $F(u, v) = u^s$ then $D(n, \chi_1, \chi_2)$ becomes the function $\sigma_s(n, \chi_1, \chi_2)$,

$$\sigma_s(n, \chi_1, \chi_2) = \sum_{d/n} \chi_1(d)\chi_2\left(\frac{n}{d}\right) d^s.$$

Here s denotes the complex variable $\sigma + i\,t$ where σ and t are real. For this function the foregoing expansion takes the form

$$\sigma_{-s}(n, \chi_1, \chi_2) = c_1^{(\lambda)}(n, \chi, \chi_1, \chi_2) \sum_{v=1}^{t} \frac{\chi(v)}{v^{s+\lambda}}$$
$$+ c_2^{(\lambda)}(n, \chi, \chi_1, \chi_2) \sum_{v=1}^{1/2\,t} \frac{\chi(v)}{(2v)^{s+\lambda}} + \ldots, t \geqq n.$$

If $\sigma + \lambda > 1$ we may let t become infinite and so obtain the following expansion:

$$\sigma_{-s}(n, \chi_1, \chi_2) = L(s + \lambda, \chi) \sum_{v=1}^{\infty} \frac{c_v^{(\lambda)}(n, \chi, \chi_1, \chi_2)}{v^{s+\lambda}},$$

where $L(s, \chi)$ denotes the L-series

$$L(s, \chi) = \sum_{a=1}^{\infty} \frac{\chi(a)}{a^s}.$$

We have also the expansion

$$\sigma_s(n, \chi_1, \chi_2) = n^s L(s + \lambda, \chi) \sum_{v=1}^{\infty} \frac{c_v^{(\lambda)}(n, \chi, \chi_2, \chi_1)}{v^{s+\lambda}}.$$

4. Putting $\lambda = 1$, $\chi_2(a) \equiv 1$ and $\chi_1(a) \equiv \chi(a)$ in the second foregoing expansion of σ_{-s} we have

$$\sigma_{-s}(n, \chi, 1) = L(s+1, \chi) \sum_{q=1}^{\infty} \frac{\chi(q) c_q(n)}{q^{s+1}}, \quad \sigma > 0.$$

From this relation it may be shown that

$$\frac{1}{m} \sum_{n=1}^{m} c_\rho(n) \sigma_{-s}(n, \chi, 1) = L(s+1, \chi) \rho^{-s-1} \chi(\rho) \varphi(\rho) + O\left(\frac{1}{m}\right), \quad \sigma > 2.$$

This general asymptotic relation is of considerable interest. Special cases of it have appeared in several papers (see Dickson's *History of the Theory of Numbers*, vol. I, pp. 291–4, 301, 322).

Putting $\chi(a) \equiv 1$, we have

$$\frac{1}{m} \sum_{n=1}^{m} c_\rho(n) \sigma_{-s}(n, 1, 1) = \frac{1}{\rho^{s+1}} \varphi(\rho) \left(1 + \frac{1}{2^{s+1}} + \frac{1}{3^{s+1}} + \dots\right) + O\left(\frac{1}{m}\right), \quad \sigma > 2.$$

If λ is an integer greater than unity and we put $s = 2\lambda - 1$, we have

$$\frac{1}{m} \sum_{n=1}^{m} c_\rho(n) \sigma_{-(2\lambda-1)}(n, 1, 1) = \frac{\varphi(\rho)}{\rho^{2\lambda}} \pi^{2\lambda} \frac{2^{2\lambda-1} B_\lambda}{(2\lambda)!} + O\left(\frac{1}{m}\right),$$

where the B's are the Bernoulli numbers $B_1 = 1/6$, $B_2 = 1/30$, ...

For the case $\rho = 1$ the last formula leads to the following theorem:

When λ is an integer greater than unity, the sum of the $(2\lambda - 1)^{\text{th}}$ powers of the reciprocals of the divisors of n is in the mean (on the average) equal to

$$\pi^{2\lambda} \frac{2^{2\lambda-1} B_\lambda}{(2\lambda)!},$$

the error term of the average for $n = 1, 2, \dots, m$ being $O(1/m)$.

When $\chi(a)$ is the principal character $\chi_0(a)$ we have $\sigma_{-s}(n, \chi_0, 1)$ equal to the sum $\gamma_{-s}(n, k)$ of the s^{th} powers of the reciprocals of those divisors of n which are prime to the modulus k. Then we have

$$\frac{1}{m} \sum_{n=1}^{m} c_\rho(n) \gamma_{-s}(n, k) = \frac{\chi_0(\rho) \varphi(\rho)}{\rho^{s+1}} \sum_q{}' \frac{1}{q^{s+1}} + O\left(\frac{1}{m}\right), \quad \sigma > 2,$$

where the prime in Σ' indicates that q ranges over the integers which are prime to k.

From this result we have the remarkable relation

$$\lim_{m=\infty} \frac{1}{m} \sum_{n=1}^{m} c_{d\vartheta}(n) \gamma_{-s}(n, k) = 0, \quad (d, k) > 1, \sigma > 2.$$

If $\chi(a)$ is replaced by the non-principal character $R(a)$ modulo 4, then from the first two equations of this section we have

$$\sigma_{-s}(n, R, 1) = \left(1 - \frac{1}{3^{s+1}} + \frac{1}{5^{s+1}} - \dots\right)\left(c_1(n) - \frac{c_3(n)}{3^{s+1}} + \frac{c_5(n)}{5^{s+1}} - \dots\right), \sigma \geqq 0,$$

$$\frac{1}{m}\sum_{n=1}^{m} c_\rho(n)\sigma_{-s}(n, R, 1) = \frac{R(\rho)\varphi(\rho)}{\rho^{s+1}}\left(1 - \frac{1}{3^{s+1}} + \frac{1}{5^{s+1}} - \dots\right) + O\left(\frac{1}{m}\right), \sigma > 2.$$

5. Incidental to the investigation of these σ-functions it was proved that

$$\sum_{q=1}^{\infty} \frac{1}{q} c_q^{(\lambda)}(n, \chi_0, \chi_1, \chi_2) = 0.$$

Taking $\chi_2(a) \equiv 1$ and $\chi_1(a) \equiv \chi_0(a)$ we find that

$$\sum_{q}{}' c_q^{(\lambda)}(n) = 0.$$

where the sum is taken for all positive integers q which are prime to a given positive integer k. The case $k = 1 = \lambda$ of this remarkable relation is due to Ramanujan (loc. cit., p. 185).

6. The expansions of σ-functions employed in this paper lead to remarkable expansions of functions relating to the representation of integers as sums of squares. We shall content ourselves with stating three particular results. Using $N[n = x^2 + y^2]$ to denote the number of ways of representing n as a sum of two squares and employing similar notations for other cases, we have

$$N[n = x^2 + y^2] = 2\pi\left\{c_1(n) + \frac{1}{5}c_5(n) + \frac{1}{9}c_9(n) + \dots\right\},$$

$$N[n = x^2 + 2y^2] = \pi\sqrt{2}\left\{c_1(n) + \frac{1}{3}c_3(n) + \frac{1}{9}c_9(n) + \frac{1}{11}c_{11}(n) + \dots\right\},$$

$$N[2^\alpha m = x^2 + y^2 + z^2 + t^2,\ m \text{ odd}] = g(\alpha)\pi^2 m\left\{c_1(m) + \frac{c_3(m)}{3^2} + \frac{c_5(m)}{5^2} + \dots\right\},$$

where $g(\alpha) = 1$ or 3 according as $\alpha = 0$ or $\alpha > 0$ and where in the second formula the subscripts appearing are those of the forms $8x + 1$ and $8x + 3$.

QUADRATIC FUNCTIONAL FORMS IN A COMPOSITE RANGE[1]

BY A. D. MICHAL AND L. S. KENNISON

DEPARTMENT OF MATHEMATICS, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated July 31, 1930

1. *Transformations of Third Kind in a Composite Range.*—Let $\bar{y}^1$, $\bar{y}^2$, ..., $\bar{y}^n$ be n independent variables and $\bar{y}^\alpha$ a real continuous function of a real variable α, defined for $a \leq \alpha \leq b$. Consider the linear functional transformation from $\bar{y}^\alpha$, $\bar{y}^1$, ..., $\bar{y}^n$ to new variables y^α, y^1, ..., y^n

$$\left.\begin{aligned} y^\alpha &= K^\alpha \bar{y}^\alpha + K^\alpha_\beta \bar{y}^\beta + K^\alpha_j \bar{y}^j \qquad (K^\alpha \neq 0) \\ y^i &= \qquad\quad K^i_\beta \bar{y}^\beta + K^i_j \bar{y}^j \end{aligned}\right\} \tag{1}$$

In (1) we assume that K^α, K^α_1, ..., K^α_n, K^1_α, ..., K^n_α are continuous functions of α, defined for $a \leq \alpha \leq b$, and that K^α_β is a continuous function of α and β, defined for $a \leq \alpha, \beta \leq b$. Here and throughout this paper we shall understand that any Greek letter used as an index can range over the closed continuous interval (a, b), while any Latin letter can take on any integral value 1, 2, ..., n. We shall use the convention of denoting Riemann integration on (a, b) by the repetition of a Greek subscript and superscript in a term except when an index is inclosed in a parenthesis. A similar convention is to hold for summation from 1 to n on the Latin indices.

The totality of such transformations (1) whose bordered Fredholm determinants do not vanish form a group G with inverses. By the bordered Fredholm determinant D is meant the following functional

$$|K^i_j| + \sum_{m=1}^{\infty} \frac{1}{m!} \Delta^{\alpha_1 \cdots \alpha_m}_{\alpha_1 \cdots \alpha_m}, \tag{2}$$

where

$$\Delta^{\alpha_1 \cdots \alpha_m}_{\beta_1 \cdots \beta_m} = \begin{vmatrix} \dfrac{K^{\alpha_1}_{\beta_1}}{K^{(\beta_1)}} & \cdots & \dfrac{K^{\alpha_1}_{\beta_m}}{K^{(\beta_m)}} & K^{\alpha_1}_1 & \cdots & K^{\alpha_1}_n \\ \vdots & & \vdots & \vdots & & \vdots \\ \dfrac{K^{\alpha_m}_{\beta_1}}{K^{(\beta_1)}} & \cdots & \dfrac{K^{\alpha_m}_{\beta_m}}{K^{(\beta_m)}} & K^{\alpha_m}_1 & \cdots & K^{\alpha_m}_n \\ \dfrac{K^1_{\beta_1}}{K^{(\beta_1)}} & \cdots & \dfrac{K^1_{\beta_m}}{K^{(\beta_m)}} & K^1_1 & \cdots & K^1_n \\ \vdots & & \vdots & \vdots & & \vdots \\ \dfrac{K^n_{\beta_1}}{K^{(\beta_1)}} & \cdots & \dfrac{K^n_{\beta_m}}{K^{(\beta_m)}} & K^n_1 & \cdots & K^n_n \end{vmatrix}.$$

We shall denote the bordered Fredholm determinant corresponding to (1) by

$$D\begin{bmatrix} \dfrac{K^\alpha_\beta}{K^{(\beta)}} & K^\alpha_j \\ \dfrac{K^i_\beta}{K^{(\beta)}} & K^i_j \end{bmatrix}$$

2. *The Form and a Functional Invariant.*—Consider a functional form with continuous coefficients

$$\left.\begin{aligned} g_{\alpha\beta}\, y^\alpha\, y^\beta + g_\alpha (y^\alpha)^2 + 2g_{\alpha i}\, y^\alpha\, y^i + g_{ij}\, y^i\, y^j; \\ g_\alpha \neq 0, \quad g_{\alpha\beta} = g_{\beta\alpha}, \quad g_{ij} = g_{ji} \end{aligned}\right\} \tag{3}$$

in the composite range $y^x, y^1, \ldots, y^n$. We assume that we are dealing with an absolute form under the group G.

As a consequence of the law of transformation of the coefficients g, it follows that the continuity of the coefficients, the non-vanishing of g_α, and the symmetry relations of $g_{\alpha\beta}$ and g_{ij} are invariant properties under G.

Since the bordered Fredholm functional of the product of two transformations of G is the product of the bordered Fredholm functionals of the transformations, it follows after some reductions that

$$D\begin{bmatrix} \dfrac{\bar{g}_{\alpha\beta}}{\bar{g}_\beta} & \bar{g}_{\alpha j} \\ \dfrac{\bar{g}_{i\beta}}{\bar{g}_\beta} & \bar{g}_{ij} \end{bmatrix} = \left(D\begin{bmatrix} \dfrac{K^\alpha_\beta}{K^{(\beta)}} & K^\alpha_j \\ \dfrac{K^i_\beta}{K^{(\beta)}} & K^i_j \end{bmatrix} \right)^2 D\begin{bmatrix} \dfrac{g_{\alpha\beta}}{g_\beta} & g_{\alpha j} \\ \dfrac{g_{i\beta}}{g_\beta} & g_{ij} \end{bmatrix}$$

under the group G. Hence we are led to the theorem.

THEOREM I. *The bordered Fredholm determinant of the coefficients of the form* (3) *is a relative functional invariant of weight two under the group G.*

$$D\begin{bmatrix} \dfrac{g_{\alpha\beta}}{\sqrt{g_\alpha g_\beta}} & \dfrac{g_{\alpha j}}{\sqrt{g_\alpha}} \\ \dfrac{g_{i\beta}}{\sqrt{g_\beta}} & g_{ij} \end{bmatrix}$$

is another form of the same invariant that exhibits explicitly the symmetric character of the original matrix.

3. *Forms Quadratic in a Function and Its Derivative.*—Let w^α be a function of α with a continuous derivative y^α. The form with continuous coefficients

$$\left.\begin{aligned} &A_{\alpha\beta}w^\alpha w^\beta + 2B_{\alpha\beta}w^\alpha y^\beta + C_{\alpha\beta}y^\alpha y^\beta + A_\alpha (w^\alpha)^2 + 2B_\alpha w^\alpha y^\alpha \\ &+ C_\alpha (y^\alpha)^2; \quad A_{\alpha\beta} = A_{\beta\alpha},\ C_{\alpha\beta} = C_{\beta\alpha} \end{aligned}\right\} \tag{5}$$

does not possess a unique expansion, so that its vanishing for all admissible w^α does not entail the identical vanisning of all the coefficients. Denoting w^β by Y and defining

$$E^\alpha = 1,\ E^\alpha_\beta = \begin{cases} 0 \text{ for } \alpha < \beta \leq b \\ 1 \text{ for } a \leq \beta \leq \alpha \end{cases}$$

we may write

$$w^\alpha = E^\alpha\, Y + E^\alpha_\beta y^\beta. \tag{6}$$

An application of the functional transformation (6) to the form (5) is instrumental in yielding the following theorem.

THEOREM II. *Any form* (5) *with continuous coefficients can be thrown over into a form of type* (3) *with* $n = 1$, $y^1 = Y$ *and with continuous coefficients given as follows*:

$$\left.\begin{aligned} g_{\alpha\beta} &= A_{\gamma\delta}E^\gamma_\alpha E^\delta_\beta + B_{\gamma\beta}E^\gamma_\alpha + B_{\gamma\alpha}E^\gamma_\beta + C_{\alpha\beta} \\ &\quad + A_\gamma E^\gamma_\alpha E^\gamma_\beta + B_\alpha E^{(\alpha)}_\beta + B_\beta E^{(\beta)}_\alpha \\ g_\alpha &= C_\alpha \\ g_{\alpha 1} &= A_{\gamma\delta}E^\gamma E^\delta_\alpha + B_{\gamma\alpha}E^\gamma + A_\gamma E^\gamma_\alpha + B_\alpha \\ g_{11} &= A_{\gamma\delta}E^\gamma E^\delta + A_\gamma E^\gamma. \end{aligned}\right\} \tag{7}$$

COROLLARY. *A necessary and sufficient condition that form* (5) *vanish for all admissible functions* w^α *is that* $g_{\alpha\beta}$, g_α, $g_{\alpha 1}$, g_{11} *defined by* (7) *all vanish identically.*

Finally the following theorem has been proved by us:

THEOREM III. *If* $'K^\alpha_\gamma = \dfrac{\partial}{\partial\alpha} K^\alpha_\gamma$ *exists, is bounded and is integrable, then the following relation holds*

$$D\begin{bmatrix} 'K^\alpha_\gamma E^\gamma_\beta & 'K^\alpha_\gamma E^\gamma \\ K^\alpha_\gamma E^\gamma_\beta & 1 + K^\alpha_\gamma E^\gamma \end{bmatrix} = D[K^\alpha_\beta],$$

where $D[K^\alpha_\beta]$ *is the well-known Fredholm determinant of* K^α_β.

[1] Presented to the American Mathematical Society, June 20, 1930. A complete account will be published in the *Trans. Am. Math. Soc.* Paragraphs 1 and 2 of the present note involve as special cases well-known algebraic theories as well as some recent results on functional forms. For the latter, see A. D. Michal, *Am. J. Math.*, 50 (1928), in particular pp. 476–480; A. D. Michal and T. S. Peterson, a forthcoming paper in the *Ann. Math.*

THE MIXED PROBLEM FOR LAPLACE'S EQUATION IN THE PLANE DISCONTINUOUS BOUNDARY VALUES

BY G. C. EVANS AND R. N. HASKELL

DEPARTMENT OF MATHEMATICS, THE RICE INSTITUTE

Communicated July 28, 1930

1. *Introduction.*—Like the Dirichlet and Neumann problems, for Laplace's equation in the plane, the mixed boundary value problem, in which the values of the function are assigned on part and the values of its normal derivative on the rest of the boundary, can be adequately discussed by means of particular regions. The region best adapted to the problem seems to be the semi-circle. With this to start with, both problem and solution may be stated in a form invariant of a conformal transformation, so as to apply to any simply connected plane region with more than one boundary point. And the main difficulties in the problem do not reappear if the arc portions of the boundary, on which function and derivative are respectively assigned, are more than two, in fact, even infinite in number; or if the simply connected region is replaced by a finitely connected, and in some cases, by an infinitely connected one.

In this paper we resume merely the results for the central problem. The boundary values, expressed in terms of limits of integrals of function and normal derivative extended over variable arcs near the boundary, are given as functions of limited variation on the boundary, and the class of functions is defined, in terms of which the solution is unique. The more usual problem is the special case where these functions of limited variation are merely the integrals of continuous functions.

Consider the semi-circle S in the upper half-plane of which the diameter is the segment $-1 \leq x \leq 1$, and let $M = (x, y) = (r, \theta)$ be an interior point of this semi-circular region. Let M' be the inverse of M in the circle, M_1 the reflection of M across the diameter, M_1' the inverse of M_1, and P a second point. Boundary values with respect to $\int u(M)d\theta$ are to be given on the semi-circumference s_1; boundary values with respect to $\int \frac{\partial u}{\partial n} ds$ are to be given on the diameter s_2.

If $g(M, P) = \log(M'P \cdot OM/MP)$ is the Green's function for the circle of unit radius, in the Dirichlet problem, it is evident that

$$G(M, P) = g(M, P) - g(M, P)$$

is the Green's function, in the Dirichlet problem, for the semi-circle, and we define

$$\mathfrak{G}(M, P) = g(M, P) + g(M_1, P)$$

as a Green's function, in the mixed problem, for the semi-circle. In fact,

$$\mathfrak{G}(M, P) = 0, \quad P \text{ on } s_1; \quad \frac{\partial \mathfrak{G}}{\partial y_p} = 0, \quad P \text{ on } s_2.$$

Green's theorems then lead to the following tentative solution of this kind of problem

$$u(M) = \frac{1}{2\pi}\int_0^{\pi}\left\{\frac{1-r^2}{1+r^2-2r\cos(\varphi-\theta)} + \frac{1-r^2}{1+r^2-2r\cos(\varphi+\theta)}\right\} dF(\varphi)$$
$$- \frac{1}{2\pi}\int_{-1}^{1} \log\frac{r^2+x_p^2-2rx_p\cos\theta}{1+r^2x_p^2-2rx_p\cos\theta}\, dH(x_p), \qquad (1)$$

where F and H are functions of bounded variation on their respective closed intervals $(0, \pi)$ and $(-1, 1)$.

2. *Properties of (1).*—We may assume without loss of generality that the discontinuities of $F(\varphi)$, $H(x)$ are regular at interior points of their intervals and that $F(0) = 0$. If we extend the definition of $F(\varphi)$ by writing $F(-\varphi) = -F(\varphi)$, $F(\varphi \pm 2\pi) = F(\varphi) \pm \{F(\pi) - F(-\pi)\}$ we make the discontinuities regular also at $\varphi = 0, \pi$ and $-\pi$, and the first integral of (1) takes the form

$$u_1(M) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1-r^2}{1+r^2-2r\cos(\varphi-\theta)}\, dF(\varphi). \qquad (2)$$

A similar but less convenient transformation may be made on $u_2(M)$, the second integral of (1). But in any case, the possible discontinuity of $H(x)$ at $x = -1$ or $x = 1$ contributes nothing to the value of the integral. Hence $H(x)$ *may be taken without loss of generality as continuous at* $x = 1$ *and* $x = -1$.

We consider the fluxes $\int \frac{\partial u}{\partial n}\, ds$, $\int \frac{\partial u_i}{\partial n}\, ds$ extended along simple rectifiable arcs $A'B'$ in S and integrals of the absolute value of $\partial u_2/\partial n$ extended over cross cuts γ, constituted by the portions in S of the level curves of the function which is harmonic in the upper half-plane except for a logarithmic infinity at $x = 0, y = 1$ and vanishes along the x axis. From lemmas of the type of the following:

LEMMA. Let $u(M)$ be harmonic in S, $\int \left|\partial u/\partial n\right| ds$ be bounded uniformly over the cross cuts γ, and $u(M)$ vanish continuously at points of the *open* interval $0 < \theta < \pi$. If $\lim \int_{A'}^{B'} \frac{\partial u}{\partial n}\, ds = 0$ as A', B' approach interior points of s_2 along orthogonal trajectories to the family γ, then $u(M) \equiv 0$ in S;

and from known properties of the logarithmic potential[1] may be deduced almost immediately the following theorem:

THEOREM 1. *A necessary and sufficient condition that* $u(M)$ *may be written in the form* (1) *is that it be the sum of two functions,* $u = u_1 + u_2$, *harmonic in the open region* S, where

$$(a) \quad \int_0^{\pi} |u_1(r, \theta)| d\theta \text{ is bounded for } r < 1,$$

$$(b) \quad \lim \int_{A'}^{B'} \frac{\partial u_1}{\partial n} ds = 0, \text{ if } A', B' \text{ approach interior points } A, B \text{ of } s_2,$$

$$(\alpha) \quad \int \left| \frac{\partial u_2}{\partial n} \right| ds \text{ is bounded uniformly when extended across the cross cuts } \gamma,$$

(β) $u_2(M)$ *takes on continuously the boundary values zero at interior points of* s_1.

We have also the following properties:

$$(c) \quad \lim_{r=1} \int_{\theta_1}^{\theta_2} u(r_1 \theta) d\theta = F(\theta_2) - F(\theta_1), \quad 0 \leq \theta_1 < \theta_2 \leq \pi,$$

$$(\gamma) \quad \lim \int_{A'}^{B'} \frac{\partial u}{\partial n} ds = H(x_B) - H(x_A), \quad -1 < x_A < x_B < 1,$$

where A', B' approach interior points A, B of s_2 along orthogonal trajectories of the family γ,
and incidentally, also

$$(d) \quad \lim_{M=P} u(M) = \left(\frac{dF}{d\theta}\right)_P \text{ in the wide sense, } P \text{ on } s_1, \text{ almost everywhere,}$$

$$(\delta) \quad \lim_{M=P} \frac{\partial u}{\partial y} = \left(\frac{dH}{dx}\right)_P \text{ in the wide sense, } P \text{ on } s_2, \text{ almost everywhere.}$$

COROLLARY. *There is one and only one function* $u(M)$ *of the class described in Theorem 1 which has the properties* (c), (γ).

This Corollary accordingly yields a unique solution of the mixed boundary value problem, for the semi-circle S. A corresponding treatment applies to the hemisphere.

3. *The General Simply Connected Region* T.—Let T be an arbitary region in the plane, simply connected and with more than a single boundary point; for convenience we take T as bounded and denote the boundary by t. The prime ends which constitute t may be regarded as ordered with respect to the conjugate of the Green's function for T with a given pole, or what amounts to the same thing, by the points of the circumference of a unit circle to which T is equivalent by a conformal transformation, or by the points of the boundary of S. Let t_1, t_2 be two portions of the boundary t which correspond to any division of the circumference

of the unit circle into two complementary arcs s_1', s_2'; the extremities and interior prime ends of t_1 and t_2 are also defined by this correspondence. Let Q, R be the extremities of t_1 in the positive order of points of t, so that R, Q are those of t_2. Let O_1 be an interior prime end of t_1 and O_2 of t_2.

The problem with respect to S has been investigated with respect to two families of curves, one with pole at the origin, the other with pole at $(x = 0, y = 1)$. The first pole may be shifted to any interior point of s_2 and the second to any interior point of s_1 and corresponding families of circles developed as level curves of harmonic functions which vanish on s_1 and s_2, respectively.

In T, therefore, we introduce two functions and their conjugates, which are themselves (like $\mathfrak{G}$) solutions of the mixed problem for special boundary values. We let $h_2(P)$ be a bounded harmonic function which takes on continuously the boundary values zero on the portion O_1R of t, except at O_1, and π on the portion QO_1 except at O_1, for which the flux $\int_{A'}^{B'} \frac{\partial h_2}{\partial n} ds$ approaches zero as A', B' approach interior prime ends of t_2, the integration being extended along a simple rectifiable arc in T. Such a function exists and is uniquely determined, as can be seen by transforming T suitably into the semi-circle S and extending the function harmonically through the entire circle.

Similarly we define $h_1(P)$, interchanging the rôles of t_1 and t_2, O_1 and O_2. We denote by $g_1(P)$ the function conjugate to $h_1(P)$ and by $g_2(P)$ the function conjugate to $h_2(P)$, the arbitrary constants being chosen so that $g_1(P)$ takes on zero values on t_1 and $g_2(P)$ on t_2.

We denote as the class (A), the class of functions $u(M)$ which have the following property: $u(M)$ is the sum of two functions, $u = u_1 + u_2$, harmonic in T such that

(a') $\int |u_1(M)| dh_1(M)$ is bounded when the integral is extended across the cross cut $g_1(M) = c_1$ as c_1 approaches zero,

(b') $\int_{A'}^{B'} \frac{\partial u_1}{\partial n} ds$ approaches zero when A', B' approach interior prime ends of t_2,

(α') $\int \left| \frac{\partial u_2}{\partial n} \right| ds$ is bounded when the integral is extended across the cross cut $g_2(M) = c_2$ as c_2 approaches zero,

(β') $u_2(M)$ vanishes continuously at interior prime ends of t_1.

Let now $F(P)$, $H(P)$ be two functions of bounded variation defined on t_1 and t_2, respectively, with regular discontinuities [as in reference 1, Chap. V]. We have the theorem

THEOREM 2. *In the class* (A) *there is one and only one function* $u(M)$ *such that*

$$\lim_{c_1=0} \int_{h_1'}^{h_1''} u(M)dh_1(M) = F_{h_1''} - F_{h_1'}, \text{ integrated along } g_1(M) = c_1,$$

where h_1' *and* h_1'' *correspond to prime ends of the closed interval* t_1,

$$\lim_{\substack{A'=A \\ B'=B}} \int_{A'}^{B'} \frac{\partial u}{\partial n}\, ds = H(B) - H(A),$$

where A, B *are interior prime ends of* t_2, *and* A', B' *have points of* A, B *as limit points of their respective sets of positions along curves* $h_2(x, y) = const.$

The class (A) is equivalent to that given by an integral formula analogous to (1) in terms of the mixed-Green's function $\mathfrak{G}$ for T.

4. *Special Cases and General Remarks.*—A function which is bounded satisfies (a') automatically, and the corresponding F is absolutely continuous (as a function of h_1), so that the boundary condition for t_1 may be replaced by stating that $u(M)$ takes on summable boundary values $f(P)$ (in the narrow sense) almost everywhere on t_1. Giving the boundary values of $\int \frac{\partial u}{\partial n}\, ds$ is equivalent to giving the boundary values, except for an arbitrary constant, of the function v conjugate to u. For lack of space, we must leave consideration of the determination of a holomorphic function by boundary values of real and imaginary parts on different arcs of the contour to another communication; suffice it to say here that the problem is uniquely solvable if the real part takes on continuous values on the closed portion t_1 of t, while the imaginary part takes on continuous boundary values $H(P)$, of bounded variation, on the open arc t_2, a problem which is a particular case of that already considered. Villat finds a particular solution of the case where u and v take on summable values on the respective arcs, with certain restrictions, but such solutions are not uniquely determined.[2]

The work of the authors of the present paper, although formulated independently during the last two years, appears as a combination of the methods of the article just referred to, and of an investigation by Lichtenstein on the mixed problem for more general types of partial differential equations of elliptic type.[3] The latter employs the method of images in constructing Green's functions, and discusses the uniqueness of the solution for continuous boundary values. Reference should be made to Zaremba's fundamental treatment of the problem in three dimensions, for sufficiently smooth surface boundaries, with continuous u and continuously given boundary values of u and $\partial u/\partial n$.[4] The continuity of $u(M)$ enters essentially into the lemma by means of which the uniqueness of the solution

is established. It is likely that Zaremba's methods can be extended. But the present investigation seems to be the only one in which the discontinuous problem is solved for its own sake, with the definition of a class of functions in which the solution is unique.[5]

In closing, let us point out that an analysis in terms of the level curves of the Dirichlet-Green function for T and its conjugate is not sufficient for the class of functions here considered, on account of the rôle of the extremeties of s_1 as critical points in S_1, or of the corresponding points as branch points if the region is the unit circle. In T there are harmonic functions, not identically zero such that $\int |u| dh$ is bounded on curves $G = \text{const.}$, such that $\int_{h_1}^{h_2} udh$ approaches zero for all portions of the closed interval t_1, while $\int_{A'}^{B'} \frac{\partial u}{\partial n} ds$ approaches zero for all portions of the open interval t_2. In S such a function is $(1 - r^2)/(1 + r^2 - 2r \cos \theta)$. The same objection, however, does not apply to an analysis in terms of the level curves of the mixed-Green's function and its conjugate.

[1] G. C. Evans, *The Logarithmic Potential*, New York, **1927**, Chaps. 3 and 4.

[2] H. Villat, "Sur la résolution de certaines équations intégrales, et sur quelques problèmes qui s'y rattachent," *Acta Math.*, **40** (101–178), see pp. 162–167.

[3] L. Lichtenstein, "Randwertaufgaben der linearen elliptischen Differentialgleichungen," *J. Math.*, **143**, 1913. See pp. 93–105.

[4] Zaremba, "Sur un problème mixte relatif à l'équation de Laplace," *Bull. Acad. Sci. Cracow*, July, **1910**.

[5] G. E. Raynor announces a contribution to the mixed boundary value problem and its unique solution at the meeting of the Am. Math. Soc. of December, 1929. See *Bull. Amer. Math. Soc.*, **36**, 1930, p. 207.